

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 11:24:22 ; Search time 24 Seconds
(without alignments)
809.130 Million cell updates/sec

Title: US-10-007-693-139
Perfect score: 3204
Sequence: 1 MHNNHHMESGESVSSNQS.....PAFIQVVLNIALSFGYLS 660

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCtUS.COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfill1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1690	52.7	361	4	US-09-556-877-299 Sequence 299, App
2	1690	52.7	361	4	US-09-620-412C-299 Sequence 299, App
3	248.5	7.8	10182	4	US-09-134-001C-1159 Sequence 3159, Ap
4	237	7.4	2137	4	US-09-134-001C-4463 Sequence 4463, Ap
5	235	7.3	2285	4	US-09-308-375-2 Sequence 2, Appli
6	219	6.8	3696	4	US-09-134-001C-5080 Sequence 5080, Ap
7	194.5	6.1	1566	2	US-08-687-956A-23 Sequence 23, Appli
8	194	6.1	1093	5	PCT-US93-03077-1 Sequence 1, Appli
9	189.5	5.9	3248	1	US-08-353-700-1 Sequence 1, Appli
10	189.5	5.9	3248	5	PCT-US95-16216-1 Sequence 1, Appli
11	186	5.8	3788	4	US-09-336-447A-76 Sequence 76, Appli
12	183.5	5.7	955	2	US-08-428-414A-3 Sequence 3, Appli
13	183.5	5.7	1561	3	US-08-894-017-23 Sequence 23, Appli
14	183.5	5.7	2101	1	US-08-466-390-4 Sequence 4, Appli
15	183.5	5.7	2101	1	US-08-470-950-4 Sequence 4, Appli
16	183.5	5.7	2101	1	US-08-467-781-4 Sequence 4, Appli
17	183.5	5.7	2101	1	US-08-195-487-4 Sequence 4, Appli
18	183.5	5.7	2101	2	US-08-483-924-4 Sequence 4, Appli
19	183.5	5.7	2101	4	US-08-483-924-4 Sequence 4, Appli
20	183.5	5.7	2101	4	US-09-452-294-1 Sequence 1, Appli
21	182.5	5.7	955	1	US-08-006-676B-1 Sequence 1, Appli
22	182.5	5.7	955	1	US-08-282-845-2 Sequence 1, Appli
23	182.5	5.7	955	5	PCT-US94-00324-1 Sequence 2, Appli
24	182	5.7	892	4	US-09-336-447A-5 Sequence 1, Appli
25	180	5.6	2482	4	US-08-328-254-6 Sequence 5, Appli
26	178.5	5.6	1354	3	US-08-685-871-2 Sequence 2, Appli
27	175	5.5	1939	4	US-09-310-187A-1 Sequence 1, Appli

28	174	5.4	2411	4	US-09-268-347-36 Sequence 36, Appli
29	173.5	5.4	1886	4	US-08-938-105-3 Sequence 3, Appli
30	172.5	5.4	873	4	US-09-336-447A-13 Sequence 13, Appli
31	172	5.4	534	4	US-09-103-664A-2 Sequence 2, Appli
32	171	5.3	676	4	US-09-336-447A-15 Sequence 15, Appli
33	170.5	5.3	889	4	US-09-134-001C-4318 Sequence 4318, Ap
34	170.5	5.3	3111	2	US-08-460-309-4 Sequence 4, Appli
35	170.5	5.3	3111	2	US-08-125-077-4 Sequence 4, Appli
36	169.5	5.3	1565	6	5352450-2 Patent No. 5352450
37	167	5.2	1073	4	US-09-541-782-6 Sequence 6, Appli
38	167	5.2	1073	4	US-09-723-820-6 Sequence 6, Appli
39	166.5	5.2	1786	4	US-08-973-462-8 Sequence 8, Appli
40	166.5	5.2	1912	1	US-08-409-995-4 Sequence 4, Appli
41	166.5	5.2	1912	3	US-08-685-467-4 Sequence 4, Appli
42	166.5	5.2	2353	4	US-09-377-155-33 Sequence 33, Appli
43	166.5	5.2	2353	4	US-08-913-942-4 Sequence 4, Appli
44	166.5	5.2	2353	4	US-09-669-974-33 Sequence 33, Appli
45	166.5	5.2	2354	4	US-09-268-347-47 Sequence 47, Appli

ALIGNMENTS

RESULT 1									
US-09-556-877-299									
Sequence 299, Application US/09556877									
Patent No. 6432916									
GENERAL INFORMATION:									
APPLICANT: Probst, Peter									
APPLICANT: Bhatia, Ajay									
APPLICANT: Skeiky, Yasir									
APPLICANT: Fling, Steve									
APPLICANT: Maisonneuve, Jeff									
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND									
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION									
FILE REFERENCE: 210121.469C5									
CURRENT APPLICATION NUMBER: US/09/556.877									
NUMBER OF SEQ ID NOS: 305									
SOFTWARE: FastSeq for Windows Version 3.0/4.0									
SEQ ID NO 299									
LENGTH: 361									
TYPE: PRT									
ORGANISM: Chlamydia									
US-09-556-877-299									
Query Match									
Best Local Similarity 97.0% Pred. No. 9e-109;									
Matches 350; Conservative 4; Mismatches 7; Indels 0; Gaps 0;									
QY	300	HTGLTDSPLVKKAEQISOAKDIOEIKRPSGSDIPYVPGSGASAGSAGALKSSNNG	359						
DB	1	HOEIDNSPLVKKAEQIQNOADQDIOITIPSGDIPYVPGSGASAGSAGALKSSNNG	60						
QY	360	RISLLDDVDNMAIATGQFPMTEQFNPNVNPATAKELQMEQULTMSSDQVAGDEL	419						
DB	61	RISLLDDVDNMAIATGQFPMTEQFNPNVNPATAKELQMEQULTMSSDQVAGDEL	120						
QY	420	PAETQIKDALQALQKPSADGLATAMGOVAFAAKVGGSGAGTAVQVMVKKQLYKTA	479						
DB	121	PAETQIKDALQALQKPSADGLATAMGOVAFAAKVGGSGAGTAVQVMVKKQLYKTA	180						
QY	480	STSSSSSTAALSDQYSATKTLNLSYSSRGVQSAISQOTANPALSRYSVSGISQGRS	539						
DB	191	STSSSSSTAALSDQYSATKTLNLSYSSRGVQSAISQOTANPALSRYSVSGISQGRS	240						
QY	540	ADASORAEITVRODQITGDVYSRLQVLDLSMTSTVSNPQANOEITMOKLTAISKARQ	599						
DB	241	ADASORAEITVRODQITGDVYSRLQVLDLSMTSTVSNPQANOEITMOKLTAISKARQ	300						
QY	600	GYPVQNSADSLOKFAQLEREFDGERSLAESQENAFKRPQAFIOQVLVNIASLFGYL	659						
DB	301	GYPVQNSADSLOKFAQLEREFDGERSLAESQENAFKRPQAFIOQVLVNIASLFGYL	360						

OY 660 S 660
Db 361 S 361

RESULT 2
US-09-620-412C-299

; Sequence 299, Application US/09620412C
; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Fling

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO: 299

; LENGTH: 361

; TYPE: PRT

; ORGANISM: Chlamydia

US-09-620-412C-299

Query Match

Best Local Similarity 97.0%; Score 1690; DB 4; Length 361;

Matches 350; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 300 HTLITSPLYKKAEOISAKOIOETKPSGSDIPYVPGSGSAASGSAAGALKSSNSG 359
Db 1 HOBIASPLVKKAEQINQAOODIQITTPSGLDIPYVPGSAAASGSAAGALKSSNSG 60
OY 360 RISLLDDVDNEMAAIATLOGFRSMIEOFNNNPATAKELOAMEAOLTAMSDOLVGADGEL 419
Db 61 RISLLDDVDNEMAAIATMGGFRSMIEOFNNNPATAKELOAMEAOLTAMSDOLVGADGEL 120
OY 420 PAETQAIKDALQALQOPSADGLATAMGOVAFAAKVGSGSAGTACTVQNNVKKLYKTAA 479
Db 121 PAETQAIKDALQALQOPSADGLATAMGOVAFAAKVGSGSAGTACTVQNNVKKLYKTAA 180
OY 480 SSS 539
Db 181 SSS 240
OY 540 ADSORAETIVDSQTLGIVYSLQVLDLMTSTVSNPOANOBEIQTASTISKAPOF 599
Db 241 ADSORAETIVDSQTLGIVYSLQVLDLMTSTVSNPOANOBEIQTASTISKAPOF 300
OY 600 GYPVONSADSLQKFAQLEREFVDEGRSLAESQENAFKOPAFIOQVLNINSLFSGYL 659
Db 301 GYPVONSADSLQKFAQLEREFVDEGRSLAESQENAFKOPAFIOQVLNINSLFSGYL 360
OY 660 S 660
Db 361 S 361

RESULT 3

US-09-134-001C-3159

; Sequence 3159, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match

Best Local Similarity 19.7%; Score 248.5; DB 4; Length 10182;

Matches 141; Conservative 128; Mismatches 256; Indels 189; Gaps 31;

OY 18 NQSSAMPILINGQIASNSFKESTKASEASPSASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 77
Db 6003 SSPTIQLNANVATQAKSNLHGDKLOHDKDSAKQOTIAQOLNLSAOK---HHEDSLDNE 6059
OY 78 SSPTD---SLSOLEASTSTVTRVAKDYDEAKN-----FDTAKSLENA 121
Db 6060 STRTQVQHDLEAQLDGLMGALKEISKDTNIVSGNNTINAEPSKKQAYDAVQAQNI 6119
OY 122 KTLAEYET-KMADMAALODMERLANS-----DPSNNHTEEVNNIKALEAOKD--- 169
Db 6120 INGTNPPTINKGNVTTATQVTKNTKDALDGRLEPAKNNANQITRLNLSNNAQDAEK 6179
OY 170 -----TIDKLNK-LVTLQONKSLTEVLTKTDSDQIPA-----INSOLETKNSADQI 217
Db 6180 NLVNSASTLEQVOQNIQOTQOQLDNMGELQSIJAKKDYKASQKYLNEEDPOIKQNDQAV 6239
OY 218 IKDLERQNTSYEAVLNTNAGEVIKASSEAGIKLQALQSIYDAGDQSOAAVLAQOQNSPD 277
Db 6240 -----QVETIINETQNPBLKAN-----IDQATQSVQNA-----EQALHGAELKLNQK 6283
OY 278 NIATK-----ELIDAETRVNE-----LKO---EHGTLTSPLYKKAEOISQA 319
Db 6284 QTSSTLEGLTDLTDPAQREKREIQNTSRRDQIKQIKQAKALNDA--MKLKEGV--A 6339
OY 320 QKDIQETKPSGSDIPYVPGSGSAASGSAAGALKSS-----NSGRISLLDDVDNEMAAI 375
Db 6340 QKQGVH---ANSQY---TNEQSAQKDAYNNALQKQAEIITNNSNPNLNAQDITNALNMT 6392
OY 376 A-----LOGFRSMIEOFNNNPATAKELOAMEAOLTAMSDOLVGADGELPAETQAIKDAL 430
Db 6393 KQADNLHGAQKLOQDKQNTTQALIGNLNLNQKRALQALQALNGAITS-----RDQV 6443
OY 431 AQALQOPSADGLATAMGOVAFAAKVGSGSAGTACTVQNNVKKLYKTAASTSSSSSYAAA 490
Db 6444 AEKLKE--AEALDEAM-----KOLEQVQVODQDISNSPF 6476
OY 491 LSDGYSAKTLNLSYESRSGVQSAISQIANPALSRVSRSIGESGRSADSQRAETI 550
Db 6477 INEDSPKQKT---YNDKIQAAKEIITNQTNSPTLTK-----QKIDATLU 6515
OY 551 ---VRD-----SQTLDVYSRLQVLDL-----MSTIVSNPOANO-----EI 585
Db 6516 QNIKDAVNNLHGDKLQAKSQDANNQNLNHDLTLEQKQNHFKRLINNATRDVYQNLQLEI 6575
OY 586 MOKLTASISKAPQFGYPVONSADSLQKFAQLEREFVDEGRSLAESQENAFK 639
Db 6576 AKQNGDMSTLHK-----VINDKQIQIOLHS-----NYINADNDKQNYDMAIKE 6619

RESULT 4

US-09-134-001C-4463

; Sequence 4463, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

;; PRIOR APPLICATION NUMBER: US 60/055,779
;; PRIOR FILING DATE: 1997-08-14
;; NUMBER OF SEQ ID NOS: 5674
;; SEQ ID NO 4463
;; LENGTH: 2137
;; TYPE: PR1
;; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 7.4%; Score 237; DB 4; Length 2137;
Best Local Similarity 19.9%; Pred. No. 1.1e-07;
Matches 136; Conservative 134; Mismatches 316; Indels 96; Gaps 23;

QY 9 ESGEVSNSQSNMPTIINGOIANSE-----TKETKASEA---SPSSSVSSWFLSS 60
DB 1079 ESDSTSTSLSESTSTSLSGSTASSTSDASTSTSESDSTSESTSLSESLTSSVSDSTAS 1138
QY 61 SAKNALISLRDAIILKNSSPTDSLQLEASTSTSTVTTRAADYDEAKSNFDYAKSGLEN 120
DB 1139 TSESASTSTSES--ESNASSTSLSGSLSTSLSDSTSTSTSTSTSTSTSTSTSTSTSTST 1190
QY 121 AKTAEIETKADLMAALQDMERLANSPSNNHTEEVNNIKALEPAQKDTIDKLNKLVTL 180
DB 1191 TSTSLSESTSA 1245
QY 181 QNOKKSILEVLTQDSDAQIPAINQLEIN-----KNSAQIILKLERONIS 227
DB 1246 STSDASTSTSTSTSDASTSL 1304
QY 228 YEAVLTNAGEYIKASSEAGIKLQALQSIIVADQSOAAVLQAOONNSPDNIATKELID 287
DB 1305 -ESISTSVSDSTASST 1357
QY 288 AAEFKVNELEKQEH-----TGLTSPFLYKAEBOISQAOKDIOEIKPGSGDIPYVPGSGA 342
DB 1358 SASSTSTSDERASTSTSLSGSTSTSL-----SDSTSTSTSDASTSTSTSTSTSTST 1409
QY 343 ASAGSAGALK-----SSNNSGRISLLDDVDNEMMAIALQGRFMSMEQNNVNNPATKELQ 399
DB 1410 SLGSLSTSVSDSTSTSTSDASTSTSTSDASTSTSTSTSTSTSTSTSTSTSTSTSTST 1454
QY 400 AMEQLTAMSDQLVGADGELPAEIQAIKDALQALQKPSADGLATVAMQVAFAAKVG-G 458
DB 1455 ISDSSTSTSDSTSD 1512
QY 459 GSAGTAGVQNNKQVLTAFSTSTSSSYAAALSDGYSAYVTLSLYSESSGVSQSAISQ 518
DB 1513 DSASTSSSESVSTSDSTSTSTSTSDASTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1569
QY 519 TANPALSRSVRSGLIESQGRSADASQRAAETIVRDSQTLGQVYSLQVLDLSMTIVENP 578
DB 1570 STSDASTSTSDSTSD 1620
QY 579 QANOELIMQKLTASISKAPQPGYPAYONSADLOKFAQLREFYDGERSLAESQENAFR 638
DB 1621 NSASTSTSLSESTSTSLSDSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1672
QY 639 KQPAFIQOVLVNIASLSESGYLS 660
DB 1673 ESTSTSTSTSVASNSSTSTSL 1694

RESULT 5
US-09-308-375-2
;; Sequence 2, Application US/09308375
;; Patent No. 6300117
;; GENERAL INFORMATION:
;; APPLICANT: Genencor International, Inc.
;; TITLE OF INVENTION: Proteases from Gram-Positive Organisms
;; FILE REFERENCE: GC394-PC1
;; CURRENT APPLICATION NUMBER: US/09/308, 375
;; PRIOR FILING DATE: 1999-05-14
;; EARLIER APPLICATION NUMBER: EP9719636.4

;; EARLIER FILING DATE: 1997-09-15
;; NUMBER OF SEQ ID NOS: 3
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 2285
;; TYPE: PR1
;; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match 7.3%; Score 235; DB 4; Length 2285;
Best Local Similarity 19.3%; Pred. No. 1.6e-07;
Matches 159; Conservative 141; Mismatches 312; Indels 210; Gaps 33;

QY 10 SGPEVSSNSQSNMPTIINGOIANSE-----TKETKASEA---SPSSSVSSWFLSS 61
DB 56 SAIDTYQKRLKSYQVYAEVETSTVIKMDGSVEKLTQOKKNGEYLQRETKIINN----- 109
QY 62 AKNALISLRDAIILKNSSPTDSLQLEASTSTSTVTTRAADYDEAKSNFDYAKSGLEN 115
DB 110 -RNTAKQEQEVKVKLQATEKLEQVQKKTQVQRLQGGPTVYQKNNRGGFDDIYTTDPK 168
QY 116 SGLENAKTLAEYETKADLMAALQDMERL-----ANSDPSNNHTEEVN 158
DB 169 TNSTSKTTTNYDQORRAIEQLKODLEKLRQGGIYTTDTTISLARKINTAQSAAQIIEALQ 228
QY 159 NIKKALEKQKPTIDKLNK-----VTLOONKSLTEVLKTTDS---ADQIPAI 203
DB 229 NRIIRLDRKSAVAANKNLKTTIELYQROAQVNVQNLTRYGSSNGSSRQAVODYLNAV 288
QY 204 NSOLEINKNSADQIILKLERONISYEAVLTNAGEYIKASSEAGIKLQALQSI---VDAG 260
DB 289 NS-LNVSTGGS--NNIRSOQLSMQFRELASNAQTAANGASSFGALQYTFMSMTYLLSG 346
QY 261 D-----OSQAAVLQAOONNSPDNIATKELIDAAETVNELEKQEHGTLTDSPL 308
DB 347 SLFGAISGLKEMYSQALIEITLMTN-----IRRVNEDPYRNELQKRSIDLGDT-L 398
QY 309 VKKA-----EEOISQAOKDIOEIKPGSGSDIPYVPGSGSASAGSAGAL 352
DB 399 SNKTTDILQMTGDFGRMGFDESELTLTQTAQVLD---NVSIDLPTDVTTLTAAMLNF 454
QY 353 K-SSNNSGRISLLDDVDNEMMAIALQ-----GFRSMI----- 384
DB 455 NIANDSISIDKLENVQNNNAVTTLDANSIRKAGSTAFGEYELNDLIGYTTAIASTT 514
QY 385 -EQPNV-----NNPATKELQAMEAQLTAMSDQLVGADGELPAEIQAIKDAL 430
DB 515 RESGNIVNSLKTIFARIGNNOSSIKALEQIGISVKTAGGEAKSA-SDLISEVAGKMDTL 573
QY 431 AQALQKPSADGLA-----TAM---QVAFPAAKVGGSGAGT-----AGYQNNV 471
DB 574 SDAQOQNTSIVAGIYQLSRFNAMNNFSIAQNAKTAANSTGSMSEQOKYADSLQARV 633
QY 472 KQLYK--TAFSSTSSSYAAALSDGYSAY-KTNSLYSESSGVSQSAISQ-----AISQTA 520
DB 634 NKIQNNFTFEFAIAMSDF---ISDGLIEFTQAAGSLNASTGVIKSVGFPLPLAAVSTA 690
QY 521 NPALS-----SVRSGLIESQGRSADASQRAAETIVRD---SQTFLG 558
DB 691 TLLSKNTRTLASSLILGTRAMQETTLATAGLEGMTAAVASRYLKLALGLLVTSTLVG 750
QY 559 DVSYRL-QVLDLSMTSTIVSNPANOELIMQKLTASISKAPQPGYPAYONSADLOKFAAQ 617
DB 751 GAFALGWALESLLISFEAKKAKDD-----FEOSQOTNWEALITTKNDSTDKLIQ 801
QY 618 L-EREFYDGERSLAESQENAFRQPAFIQOVLVNIASLSESGYLS 658
DB 802 YKQLQKVESRSLSNDEQEVLTQ---VTQQLAQGFPPALVKCY 840

RESULT 6
US-09-134-001C-5080
;; Sequence 5080, Application US/09134001C

Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5080
LENGTH: 3696
TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 6.8%; Score 219; DB 4; Length 3696;
Best Local Similarity 20.3%; Pred. No. 4,1e-06;
Matches 142; Conservative 125; Mismatches 307; Indels 126; Gaps 28;

QY 3 HHHHHHSSPPES-----VSSMOSSMNPILINGQIASNSETKESTKASESPASSSVSS 55
DB 1084 HTNVNVQKRPQAKQALIAKTNEKQSAI NSDNESTIEKQAKIQS--LNDAKNLADBOITQ 1141
QY 56 WSEFLSSAKNAL-----IS-----LRDAIILNK-----NSSP-----TDS 83
DB 1142 AASNOVNDALNIGINISIKIQTNTFKQOARQOVNOKFOEKAELNSPTPHATQDEKQDA 1201
QY 84 LSOLEASTSTSTVTVAAKDYDEAKS--NPDVA-KSGLE-----NAKTLAEYETKMD 133
DB 1202 LTRITQAKET-----ALNDINQAOQNVNDPALTSGIONIQNTQVNVKRRQKQAKTTIND 1255
QY 134 LMAADQMERLANSPPSNHTEV--NNIKALEAQKDTIDKLKLVTLQNOKSLTEVLK 192
DB 1256 IVO--QHKOSIQONDDATTEEEKEVANNLVNA--SOQNVISKIDN--ATTNNQIDGI----- 1305
QY 193 TTDSADQIPAINSOLEIKNSADQILKLEPONISEAVLTNAGEYIKASSEGIRLGOA 252
DB 1306 VSDGRGSIINATIPDTISIKRNAKNDIDIKAKDKIKIQTIRINDATDELIQ--EANKRIEER 1362
QY 253 LOSIVDAGDQSOAAVLQAOONNSPDNIAATKELIDAEETKVNELKQEHGTGLDPSLVKKA 312
DB 1363 K--IEAKDNIOKNSRTQOVNEAKTNGINKIENTPTATTVKSEARQ-----AVQNNKA 1411
QY 313 EBDISQAQADIQEIKRPSGSDIPYVPSGSAASAGSAGAKSSNNSGRISLLIDVDNEM 372
DB 1412 NEDQINH---IQNTPDATNEEKQEAIRNVSABELARVOAQIMEHTQGVKTIKDDAITSL 1467
QY 373 AAIALQGFMSIEQFNVNVPATAKELQAMEAOULTAMS DQLVGDELPAETIOAIKALAQ 432
DB 1468 SRINMO-----VVEKESARN-----AIEQKTIQOT--QFINNDNATDEEEKVANNLVI 1514
QY 433 ALKQPSADSLATAMQVAAAPAAKVGSGSAGTAGTVOMNKKLYKTAFSSSTSSSYAALS 492
DB 1515 ATKQKSLDININSLSNNVDENKAVAG-----INEIANVLPATAVKSKAKKDIDQKLA 1566
QY 493 DQYSATKTLNLSYSESRGVSQASISOTANPALSRVSRSGIESQGSADASQAPAAETIYR 552
DB 1567 QOIINOIQTOTATTEKEKAIOALANOKSNEA-----RTAIONEHSNNGVAAQAKSNGIHE 1620
QY 553 DSGTGLDQVSRLOVNDLSL-----MSTIVSNQAOANCEIMOKLTASISKAPQFGYPAYO 605
DB 1621 IELVMPDAAKKSADAKOSIDNKKNESNTINTTTPDATDEE--KQALDKLAKIADAGYNKVD 1679
QY 606 NSADSLQKFAAOLEREFDVDEKSLAESQENAFKKQAPAFIO 645
DB 1680 QAOQTNOQ--VSDAKTEAID--TITNIQANVAKKPSARVE 1714

RESULT 7

US-08-687-956A-23
Sequence 23, Application US/08687/956A
Patent No. 5861157
GENERAL INFORMATION:
APPLICANT: BURNIE, JAMES P
APPLICANT: MATTHEWS, RUTH C
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
TITLE OF INVENTION: INFECTIONS DUE TO STREPTOCOCCI AND ENTEROCOCCI
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO, LLP
STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH
STREET: FLOOR
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,956A
FILING DATE: 29-JUL-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9401689, 6
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 50885/222892
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/861-3000
TELEFAX: 202/822-0944
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Streptococcus sobrinus
US-08-687-956A-23

Query Match 6.1%; Score 194.5; DB 2; Length 1566;
Best Local Similarity 21.0%; Pred. No. 5.8e-05;
Matches 157; Conservative 116; Mismatches 281; Indels 195; Gaps 34;

QY 13 ESYSSMOSSMNPIL-----NGQIASNSETKESTKASESPS-----ASSSVSSWSE 58
DB 40 ESTSTSGVNTAVVGTETGNPATNLPRKQDNPSQAFESQAOAGKGTGAMVDVSTSEL 99
QY 59 LSSAKNA-----LISLBDALINKNSPT--DSLQGLEASTSTSTVTRA-----AKDVE-- 106
DB 100 DEAKSAQAGVIVSODAVVDGVTGVEDPSEANQKETEIRKDYSKQAAADIQKTTEDYKAAV 159
QY 107 AKSNFPTAKSGLENATKTLAEYETKMDLMAALQDMERLANSPPSNHTEEV----- 157
DB 160 ARNQATETDITQENAKKQAYE-----QDLAANKRAEVRITNENAKRADYEAUKLAQYQKDL 216
QY 158 -----NNIKALEAQKDTIDKLKLVTLQNON-----KSLTEVL--KTDSADQIPAINS 205
DB 217 AAVQOANNDSQAAVYAAKBAVDK--ELARVOAANAARKEVEBALANFTTKNQIKAEANA 274
QY 206 QLEINKNSADQILKLERONISEAVLTNAGEYIKASSEAGIKGLGALOSIYDAGQOQA 265
DB 275 AIO-QRNA--QAKADYEAUKLAQYEKDL-----AAOSGNATNEDADYQAKKAAYEOEL 323


```

: TITLE OF INVENTION: AND METHODS OF USE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
: STREET: 1601 MARKET STREET, SUITE 720
: CITY: PHILADELPHIA
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/353,700
: FILING DATE: 09-DEC-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: REED, JANET E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3248 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: HUMAN
: US-08-353-700-1

Query Match          5.9%; Score 189.5; DB 1; Length 3248;
Best Local Similarity 20.6%; Pred. No. 0.00036;
Matches 164; Conservative 126; Mismatches 292; Indels 213; Gaps 36;

QY 13 EYSSNQSSMPTINGQINSESTKASEASPSASSVSSWFLSSAKNALISLRDA 72
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 949 ETLSTLEKKEMSSIIIS-LNKREIELTQENGLKEINAS-----LNQEKMWLQKSES 999
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 IINKNSPTDSLSQ-----LEASTSSTVTRYAANDY-----DE 106
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1000 FANYIDERKESISELSDQYKQEKLLILOKCEETGNAYEDLSQYKKAQKESKLECLNE 1059
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 107 AKSNEDTAKSGLEN-----AKTLAEYETKMA-----DLMALQDMERLANS--DPSN 151
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1060 CTSICLENKRNLEQJKEAFKAEHQEFLTKLAFAEERNQNLMELETVQALREMTDNQ 1119
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 NITEEYNNIKKALEAKQITIDKLNKLYT-LQNONKSLTEVLKTTDSADOI---PAINS-- 205
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1120 NKSSEAGLKOEIIMTKQKNQKQEVNDILOENQIMKVMKTKHCQNLSESPINRSYK 1179
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 206 -----QLEINKNSAQDIINDLERQNTSYEAVLTNAGEVYKASSEAGIKIG 250
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1180 ERESERNQCNKRPQMDLEVEKESLID---SYNAQVQLEAMLRNKLKQESKEKECLO 1235
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 251 QALQSTIVDAGDQQAVALQAOQNNSPNDIATKEL-IDAAE----- 290
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1236 HELQITRGLDLETSNLODMQSOE-----ISGLKDCELDAEKKYISGPHLETSQNNAHL 1289
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 291 -----TKVNEK-----QEHFTGLTDS-----PLVKAEEQISQAQKDIOEI 326
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1290 QCSLQITMKNKLELEKICILQAEKELYTELINDSRSECTITATRKMAEEVG--KLINVEY 1346
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 327 KSGSDIPIVG-----PSSGASASAGS-----AAGALKSSNNNSGRLISLLDDVDNDEMMA 374
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1347 KILNDGSLHGLVELVDIGGEGEPQNPQHPVSLAPLDESNSYEHLLTIDREVMQHFAB 1406
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 375 -----IALQGFSSMTIEQFNVNPNPATAKELQ-----AMEAQLTAMSDQLVGADDELPAEIQ- 424
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1407 LQKFLSLQSEKHILHDQCNQSSKMSKSELQTYVDSIKAEMLVLTNRNFGDVLKEMQL 1466
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 AIKDALAQAQLKQ---PSAGLATAMQVAFAPAA---KVGGS-----AGTAGVQANVQ 473
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1467 GLEEGIVPSSLSSCVPDSSSL-SSLGSSSFYRLLEQTDMSLISNLEAVSANQCSVDE 1525
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 474 LYKTAFTSSSTSSSAYAAALDGYSAYKTLNSLYSESRGVSQASISQANPALRSV--SRS 531
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1526 VECSSLQITVYDSLKAEMLVLTNRNFGDVLKEMQLGLEGLV-----PSSLSSCVPDSS 1581
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 532 GIESQGRSADASQRAETIVRDSQTLGDVYSRLOVLDLSMTIVSNPQANQ----- 582
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1582 SLSSLDSSSFYRLLEQTDMSLISNLEAVSANQCSVDEPCS 1625
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 583 -----EETIMOKLTASISKARQFGYPANQNSAD-----SLQFPAOLE-----REFPDGR 627
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1626 SLQENMLTRKEEPS--APAKGVEELSLCEVYRSLERLEKMEQGIKMKKEIQELQ 1682
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 628 SL-AESQE-NAPRKQ 640
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1683 LLSERQELDCLRKQ 1697
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
PCT-US95-16216-1
: Sequence 1, Application PC/TUS9516216
: GENERAL INFORMATION:
: APPLICANT: Yen, Timothy J.
: APPLICANT: Ratner, Jerome B.
: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16216
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/353,700
: FILING DATE: 09-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3248 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US95-16216-1

Query Match          5.9%; Score 189.5; DB 5; Length 3248;
Best Local Similarity 20.6%; Pred. No. 0.00036;
Matches 164; Conservative 126; Mismatches 292; Indels 213; Gaps 36;

```


GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TITLE OF INVENTION: LEISHMANIASIS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,414A
FILING DATE: 21-APR-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Kadlecsek, Ann T.
REGISTRATION NUMBER: 39,244
REFERENCE/DOCKET NUMBER: 210121.407
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERRY
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 955 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-428-414A-3

Query Match 5.7%; Score 183.5; DB 2; Length 955;
Best Local Similarity 20.8%; Pred. No. 0.00017;
Matches 155; Conservative 126; Mismatches 258; Indels 207; Gaps 36;

QY 10 SCGESVSSNOSSNPIINOIASNSETKESTK-----ASASASASSSVS 54
DB 292 AGSERAAQSO-----VEQO-----OFKEATHINLSLTGLGRYIDVLADMATGAKAQS 340
QY 55 SWSFSLSAKNAALSLDPAITINKNS-----SPT-----DSLQLEASTSTVTNRVA- 100
DB 341 VAFPRDSKLTFT--LKDS--LGNSKTFMTATVSPSALNTEETLSTLRYSRANDIVNVAQ 397
QY 101 -----AKDYDEAKSNEDTAKSGLENA-----KTIAEYETKMADLMAALQDMERLA 145
DB 398 VNEDPRARRIRELEQMEDRMQAGGDPAVYSELKKLALLESEAKRAADQALER-- 455
QY 146 NSPSPNNHTEVNNIKALEAQDITDKLNKLVTLONONKSLTEVLTSDAQIPAIN 205
DB 456 ---EREHNOVOERLLRATEAEKSELE--SRAALOE-----EMTATRQADKMAALNL 503
QY 206 QL-EINKNSADQITKDERONISYEAVLTNAGEVIRKASSPAGIKLQALQSIYDAGDSQ 284
DB 504 RLKEBQARKERELKEMAKK-----DAALSIVRRRLDA-----EIASERKLE 546
QY 265 AAVLQAOONNSPDNIATRELDAETKYNELQ-----EHTGLTDSPIVKAEE----- 314
DB 547 STVAQLEREQEREVA-----IDALQTHQKIQLEALESSRTAERHQLLOLQTLQDSER 601
QY 315 -QISQA-----QKDIQETK-----PSSGDIPIVGPSSASASAGALSSNNSGR 360
DB 602 TOLSVVTRERELTRDORLOYEYGETELARVALCAQEMERHYAAVFIQOT-----L 656
QY 361 ISLLLDVDNEMAILOGFRSMIEQFNUNNPATAKELOMEA-----QLTAMS 409
DB 657 LELATEMED-----ADRE-RALAEK---DEAAALAEIDMAASTSONARESCERLTSL 705

QY 410 DOLVGAD---GELPAEIOAIKDA-----LAQALKQPSADGLATAMGQVAPA 452
DB 706 OQLRESEERAAELASQLENTAAKSSAEQDRENTRATLEQQLRESPARAEELASQLENTA 765
QY 453 AAKVGG--GSAGTACTIVONVQOLYKTA--FSSSTSSSTAALIS---DQYSAKTTLSLX 505
DB 766 AAKMSAEQDRENTRATLEQQLRSEERAAELASQLESTTAAKMSAEQDRESTRATLEQQL 825
QY 506 SESR-----SGVSAISOTANPALRSVRSRGIESQGRADASQRAAETIVRDSQTLG 558
DB 826 ROSEERAAELASQLESTTAAKMSAEQDRESTRATLEQQLR--ESERAAE----- 873
QY 559 DVYSRLQVDSIMSTIVSNPOANQEE-----IMQKLITASISAPQGYPAYONSADSLQ 612
DB 874 -----LASQLESTTAAKMSAEQDRESTRATLEQQLROSEERAAEL--ASQLENTA 923
QY 613 KFAQOLEREVDG--ERSLAESOEN 636
DB 924 KSSAEQDRENTRAALQQLNDSEERA 949

RESULT 13
US-08-894-017-23
Sequence 23, Application US/08894017
Patent No. 6024958

GENERAL INFORMATION:
APPLICANT: Lehner, Thomas
TITLE OF INVENTION: POLYPEPTIDE FRAGMENTS CAPABLE
OF COMPETITION WITH STREPTOCOCCUS MUTANS ANTIGEN I/II
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW, suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,017
FILING DATE: 31-JUL-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB96/00207
FILING DATE: 31-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 25150-20667.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-1500
TELEFAX: 202-822-0168
TELEX:
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1561 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-894-017-23

Query Match 5.7%; Score 183.5; DB 3; Length 1561;
Best Local Similarity 19.4%; Pred. No. 0.00033;
Matches 116; Conservative 110; Mismatches 270; Indels 101; Gaps 20;

QY 13 ESYVSSNOSSNPIINOIASNSETKSTKASPSASSSVSWFLSNAKNAALISLRDA 72

Db 34 OKFADETTTTSDVTKVGTOTGNPATNLPDAGSAS-----KQAEQSQTLEKRO 84
QY 73 ILNKSSPTDLSLEAFTSTSTVTRVAKODYDEAKSNFDTAKSLENAKTLAEVETKMA 132
Db 85 MHTTEVTKTDDQAKAKSAGVNV--QDADVKKGTVKTAEAVQETELREYTKQA 142
QY 133 DLMAALQMERLANSPSN--NHTEEVNNIKKALEAQKDTIDKLKLTVLQONKSLTEV 190
Db 143 -----EDIKTTDQYKSDVAHAHEAVAKIKAKNATKEQYCK--DMA-----HKAVER 190
QY 191 LKTTDSADQIPAINSOLEINKNSADQIIKLERONISYEAVLTNAGEVIKASSGIRKLG 250
Db 191 INAAHAAK--TAYEAKLAQYQADLAAYOKTNAANAQASYKALAAVQAEIKRVQENAAK 249
QY 251 QALOSIVAGQSOQAIVLQAOQONSPDNIATKELIDAEIKRVNLEKDEHGTIDSPLYK 310
Db 250 AAYDTAVANNAKNTETIAAANEIRKRNAATKAAY-----ETKLAQYQAEIKRVQENAAK 305
QY 311 KAEQD--ISOAKDIOEIKPSGSDPIVPGSGSASAGSAGALKSSNNSGRISILLDV 368
Db 306 EADYQAKLTAYOTELARQOKANAD-----AKAAYEAANNAKNA----- 346
QY 369 DNEMAIALOGFRSMIEGFVNNNPATAK--ELQAMEAQLTANSDDLVCADGELPAETQAIK 427
Db 347 -----ALTAEHTAIKQRENAKATYEAALQYEAADLAAYKAKANAHEADYQAKLTAYQ 399
QY 428 DALQALQKQPSADGLATMGQVAFPAKAVGSGSAGTACTVQ-----MNVKOLY--KT 477
Db 400 TELARVOK-----ANADAKAAVEAAYANNAANNAALTAENTAIKKRNAPAKADYEAKL 452
QY 478 AFSSTSSSYAAALSDGYSAVKTLSLVSSESRSQSAISOTANPALSRVSRSIEGSG 537
Db 453 AKYQADLAKYQKDLAD-----YPVKIKAYEDECASIKALAE-----LKHKNEDG----- 498
QY 538 RSADASQRAAETIVD-----SQTLDVYSRLQVLDLSLMTSTVSNPOANOEITQ 587
Db 499 ---NLTEPSQNLVYDLEPNANLSLTGTGKFLKASAVDARFSKSKAKYDQ-KILQ 551

RESULT 14
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESO, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MFP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4

Query Match 5.7%: Score 183.5; DB 1; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

Db 29 QIASSEETSESTKASEASPSASSSVSSWFLSSA---KNALISRLDALINKSSPTDLS 85
QY 338 QDALNELTEHSHSKATQEWLEKQAOLE--KELSAALQDKCKLEKNELQGLSLGLEHLS 395
Db 86 QL-----EASTSTVTRVAKYDEAKSNFDTAKSLENAK 122
QY 396 QLODNPQEKGEVLGDVLTQLETLKQEAATLANNLTOL-----QARVEMLETERGOQEK 449
Db 123 TLAE---YETKMAIDLMALODME--RLANSDPNNHTEEVNNIKKA--LEAO-----KDTIDK 173
QY 450 LLAEGRHFEERKQOLSLITDQSSISMLQAKBELBOASAHGRLTAQVASLTSELTY 509
Db 174 LNKLTVLQNO-----NKSLEVLKTTDSAD-----QIPAINSOLEINKNSADQ 216
QY 510 LNAITQGDQDELQGLKQQAKEKQAOQLAOTLQOQEOASGLRHQVQDLSLSLKQKQOLKE 569
Db 217 IIKLERONISYEAVLTNAGEVIKAS--SEAGIK-----LQALOSTVD 258
QY 570 VAKEQATRQDHAQOLATAAEERASLERDQALQLELEKEKAKEKLEILQOOLVANE 629
QY 259 AGDSQAANVLDAQONNSPDNIATKELIDAEIKRVNLEKDEHGTIDSPLYKKAEOIS- 317
Db 630 ARDSAQSVTQAOERKA--ELSRKVELOACVETARQOEHEAOVALELEQLRSEQOKAT 688
QY 318 -----QAKDIOEIKPSGSDPIVPGSGSASAGS-----AAGALKS-----S 355
Db 689 EKERYAQEKDQLOLOLAKES-----LKYTGSLSEERKRAADLEEQOKCIS 737
QY 356 NNSGRISILLD-----DVNEMAIALOGFRSMIEGFVNNNPATAKEL-----QAMEA 403
Db 738 ELKAETRSILVBOHREKRELEEBERA--GRKGLIARLLQLGHAHQAEFVLRLRELAEMAA 795
QY 404 QLTAMS--DQVYGADGELPAETQAIKDALQALQPSADG-----LAT-----ANGQ 448
Db 796 QHTAESECEOLY-----KEVAAMRDQYEDSQOEEAQYGMFOBQULTLKECEKARQ 848
QY 449 VAFPAKAVG-----GSAGTACTVOMNVKOLYKTTAFSSTSSSY 487
Db 849 LOEAKERYAGIESHSELDISQOKKLAEHLNMLARALQOQVEKEVRACKLADLSTLOEK 908
QY 488 AAALSDGYSAVKTLSLVSSESRSQVSAISOTANPALSRVSRSRG-----IESGRSA 540
Db 909 MAATSKYVARLETL-----VRKAGQO--QETASRELVKEPRAGDQRPVLEFGQGRQF 960
QY 541 DASQRAETIVRDSOTLGDVYSRLQVLDLSMT--IYSNPANOEIMOKLTASTISARQ 598
Db 961 CSTQAAILOAMEREAQOMONTELERLA--ALMESGQOQOQOQOQOQOQOQOQOQOQOQO 1016
QY 599 FGYPVQVNSASLOKFAOQLREFYVDGERSLAEQENAFRRQP---AFIQOVLVN 650
Db 1017 -----QADLAEKARAELEM-----RLQNNALNQRVRFATLQDALAH 1054

RESULT 15
US-08-470-950-4
; Sequence 4, Application US/08470950
; Patent No. 5696439
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX

```

: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TESTA, HURWITZ & THIBEAULT
: STREET: 125 HIGH STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02110
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/470,950
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESQ, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MTP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 248-7000
: TELEFAX: (617) 248-7100
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-470-950-4

```

```

Query Match      5.7%; Score 183.5; DB 1; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

```

```

QY 29 QIASNETKESTKASEPASPSSVSSWFLSA---KNALISLRDAIILNKSSPPTSLS 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 QDALNLTETHEHSKATQEMLEKQKQLE--KELSAALQDKCLEKEKNELGKLSQLEHLS 395
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 QL-----EASTSTSTVTVAAKDYDEAKSNFTASGLENAK 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 396 QLODNPPEKEGVLGVLOLETLKQEAATLAANNLTL-----QARVEMLETERGOQEAR 449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 TLAE---YETKMDLMAALQDME--RLANSDDPSNNHTEEVNIIKA--LEAQ---KDTIDK 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 450 LLAERGHFEERKQOSSLITDLOSSISNLSQAKEELEQASQAAGARLTAAVASTLTSLTT 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 174 LNKLVTLQNO-----NKSLEVLKTTDSAD---QIPAINSOLEINKNSADQ 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 510 LNAITQQQDQDELQAKQAKQKQALQTLQOQEQASQGLRHQVEQLSSLSKQKEQOLKE 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 217 IITDLERQNSTYFAVLNNGEVIKAS---SEAGIK-----LGAALQSTVD 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 570 VAKQETATRODHQAQALATAAEEREASLRERDAALKQLELEKKAALKLETLQOLOVANE 629
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 259 AGDQSOAAVLQAOQNNSPDNIATKELIDAETKRVNELKQEHGTGTPSPVYKKAERQIS- 317
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 630 ARBSAQTSTYTAQAREKA--ELSKRVEBELQACVETARQDQEHQAQVALEQLQLSSEQKAT 688
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 318 -----QAQKDIOETKPSGSDIPYVPGSSAASAGS-----AAGALKS-----S 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 689 EKEEVAQEKQLOLQLOALKEKES-----LKVTKGSLSEBKRRADALDEQRCIS 737
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 NNGGRISLID-----DVDNEMALALOGFSMTIQFVNVNPAATKEL-----QAMEA 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 738 ELKAETRSILVEQHKRREKKELEERA--GRKGLEARLQLEGAHQAELEVLRLRELAEMAA 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 404 QLTAMS--DQLVGADGELPAEIQAIKDALAQALKOPSADG-----LAT-----AMQ 448
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 796 QHTAESCEBQLV-----KEVAAMRDGYDSQDEEQAQYGAAMQEOBQMLTKKECEKARQ 848
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 449 VAFPAAKVGG-----GSAGTAGTVQMVVKOLYKTAFTSSTSSSY 487

```

```

DB 849 LQFAKEKVAIGISHSELQISROONKLAELHANLARALQOQKEVRAOKLADDLSTLOEK 908
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 488 AALSDGYSAVYTLNLSYESRSGVQSAISOTANPALSSVSRSG-----IESGRSA 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 909 MAITSKEVARLETL-----YRKAGEQ---QETASRELYKEPARAGDRQEMLEEQGRQF 960
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 541 DASORAETIVRDSQTLGVYSRLQVLSLMST--IVSNPQANQOEIMOKLTASISKAPQ 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 961 CSTQALQAMERAREAQMGNELELTRA--ALMESQGGQOQERQOQREVARLTQERGERA-- 1016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 599 FGTPAVQNSADSLQKFAQLEREFEVDGERSLASEQENARKQD---AFTQOVLVN 650
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1017 -----QADLALKEKARAELEM-----RLQNALNEQREVERPATLQELAH 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 16

```

US-08-467-781-4
: Sequence 4, Application US/08467781
: Patent No. 5780596
: GENERAL INFORMATION:
: APPLICANT: TOKKATLY, GARY
: APPLICANT: LIDGARD, GRAHAM P
: TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
: TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TESTA, HURWITZ & THIBEAULT
: STREET: 125 HIGH STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02110

```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/467,781
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 424

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESQ, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MTP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 248-7000
: TELEFAX: (617) 248-7100

```

```

: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-467-781-4

```

```

Query Match      5.7%; Score 183.5; DB 1; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

```

```

QY 29 QIASNETKESTKASEPASPSSVSSWFLSA---KNALISLRDAIILNKSSPPTSLS 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 QDALNLTETHEHSKATQEMLEKQKQLE--KELSAALQDKCLEKEKNELGKLSQLEHLS 395
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 QL-----EASTSTSTVTVAAKDYDEAKSNFTASGLENAK 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 396 QLODNPPEKEGVLGVLOLETLKQEAATLAANNLTL-----QARVEMLETERGOQEAR 449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 TLAE---YETKMDLMAALQDME--RLANSDDPSNNHTEEVNIIKA--LEAQ---KDTIDK 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 450 LLAERGHFEERKQOSSLITDLOSSISNLSQAKEELEQASQAAGARLTAAVASTLTSLTT 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```
GENERAL INFORMATION:
APPLICANT: TOURKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/483,924
APPLICATION NUMBER: US/08/483,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, ESQ. EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-924-4

Query Match          5.7%; Score 183.5; DB 2; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

QY 29 QIASNSETKSTKASPSASSVSSWSFLSA---KNALISLRDALINKNSPTDIS 85
DB 338 QDALNLTEHSHKATQEWLEKQAQLE--KLSAALQDKCKLEKKNIILOGKLSQLEHLS 395
QY 86 QL-----EASTSTSTVTTRVAAKDYDEAKSNFDTAKSGLENK 122
DB 396 QLODNPQKEGYLDVLOLETLKQEAATLAANNTP-----QARVEMLETERGQOEAK 449
QY 123 TLAE---YETKADLMAALQDM--RLANSDPNSNHTTEEVNNIKKA--LEAQ---KDTIDK 173
DB 450 LLAERGHFEEKQQLSSLTIDQSSISNLSQAKEELEQASQAHGALITQAVASLTSELT 509
QY 174 LNKLVTLQNO-----NKSILEVLTQTSAD-----QIPALINSOLEINKNSADQ 216
DB 510 LNAITQQOQOELAGLKQAKKEKQAQLAQTLQOQEQASQGLRHQVEQLSSLSKQKEQOLKE 569
QY 217 IIKDLERONISYEAVLTNAGEVIKAS---SEAGIK-----LQALQOSIYD 258
DB 570 VAEKQATKQDHQQLATAAEEREASLRERDAALQOLEALEKKAAKLEITLQOQLOVANE 629
QY 259 AGDQSAAVLQAOQNNSPNIAATKELIDAETKVNLEKQEHGTLTDSPLVYKKAEOIS- 317
DB 630 ARNSAQTSTVQAQOREKA--ELSKRVEBELQACVETARQEHQAQVAVLELQILRSBQOKAT 688
QY 318 -----QAQKDIQETIKPSGSDIPYVPGSSAASAGS-----AAGALKS-----S 355
DB 689 EKERVAAQERQDQLOEQALAKES-----LKYTKGSLSEEEKRRADALEEQORCIS 737
QY 356 NNSGRSLILD-----QVDNEMAAIALOGFRSMIEQFVNPNPATKEL-----QAMEA 403
DB 738 ELKAEITRSLVEQHKRERKELEBERA--GRKGLEARLLQGLEAHQAEITVLRRELAERMAA 795
```

```
QY 404 QLTAMS--DQYVAGDELPAEIOATKDALAQALKQPSANG-----LAT-----AMQO 448
DB 796 QHTAESECEQLV-----KEVAAMRDGYEDSQOEQAQYGAHQFQELMTLKECEKARE 848
QY 449 VAEPAKVGG-----GSAGACTIVOMVVKOLYTAFASTSTSSSY 487
DB 849 LQFAKKEKVAQIESHSELQISROQNKLAELHANLARLQOQVEYVAQKALADDLSLTQEK 908
QY 488 AALSDGYSAYKTLNLSYSESRGVQSAISQTPANPALSRSVSRSG-----IESQGRSA 540
DB 909 MATSKVARELETL-----VRKAGEQ---QETASRELVEKPARAGDRQEWLEEQGROF 960
QY 541 DASQRAEITIVRDSQTLGDIYVSRLOYVLDLSMT--IVSNPQANQEBIMQKLTASISKAQO 598
DB 961 CSTQAALQAMEREAQMGNELETLRA--ALMSQGGQOQEEHQREVARLTQERGERA-- 1016
QY 599 FGYPVQVNSADSLQKPAQLEREFYVGERSLAESQENAPRKOP---AFITQOVLVN 650
DB 1017 -----QADLALAEKARALEM-----RLQNALNEQRYVERATIQEALAH 1054

RESULT 19
US-09-452-294-1
Sequence 1, Application US/09452294
Patent No. 6287790
GENERAL INFORMATION:
APPLICANT: Lellievre, Sophie
APPLICANT: Bissell, Mina
TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
TITLE OF INVENTION: THERAPH AND DETECTION OF PROLIFERATIVE AND
FILE REFERENCE: IB-1454- Sequence Submittal
Patent No. 6287790
CURRENT APPLICATION NUMBER: US/09/452,294
CURRENT FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: 60/110,420
PRIOR FILING DATE: 1998-11-30
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 2101
TYPE: PRT
ORGANISM: Homo sapiens
US-09-452-294-1

Query Match          5.7%; Score 183.5; DB 4; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

QY 29 QIASNSETKSTKASPSASSVSSWSFLSA---KNALISLRDALINKNSPTDIS 85
DB 338 QDALNLTEHSHKATQEWLEKQAQLE--KLSAALQDKCKLEKKNIILOGKLSQLEHLS 395
QY 86 QL-----EASTSTSTVTTRVAAKDYDEAKSNFDTAKSGLENK 122
DB 396 QLODNPQKEGYLDVLOLETLKQEAATLAANNTP-----QARVEMLETERGQOEAK 449
QY 123 TLAE---YETKADLMAALQDM--RLANSDPNSNHTTEEVNNIKKA--LEAQ---KDTIDK 173
DB 450 LLAERGHFEEKQQLSSLTIDQSSISNLSQAKEELEQASQAHGALITQAVASLTSELT 509
QY 174 LNKLVTLQNO-----NKSILEVLTQTSAD-----QIPALINSOLEINKNSADQ 216
DB 510 LNAITQQOQOELAGLKQAKKEKQAQLAQTLQOQEQASQGLRHQVEQLSSLSKQKEQOLKE 569
QY 217 IIKDLERONISYEAVLTNAGEVIKAS---SEAGIK-----LQALQOSIYD 258
DB 570 VAEKQATKQDHQQLATAAEEREASLRERDAALQOLEALEKKAAKLEITLQOQLOVANE 629
QY 259 AGDQSAAVLQAOQNNSPNIAATKELIDAETKVNLEKQEHGTLTDSPLVYKKAEOIS- 317
DB 630 ARNSAQTSTVQAQOREKA--ELSKRVEBELQACVETARQEHQAQVAVLELQILRSBQOKAT 688
```


APPLICATION NUMBER: US/08/006,676B
FILING DATE: 15-JAN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oster, Jeffrey B.
REGISTRATION NUMBER: 32,585
REFERENCE/DOCKET NUMBER: REED-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 232 7845
TELEFAX: (206) 236 0205
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 955 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-006-676B-1

Query Match 5.7%; Score 182.5; DB 1; Length 955;
Best Local Similarity 21.5%; Pred. No. 0.00019;
Matches 158; Conservative 122; Mismatches 269; Indels 187; Gaps 36;

QY 10 SGPESVSSQSSMNPINQIANSSEKSTK-----ASEASPSASSVS 54
DB 292 AGSERVAQSQ-----VEGQ-----QFKETHINISLTLGRVIDYLDAMATKGAQAQYS 340
QY 55 SMSFLSSAKNALISLDAIILNKNS-----SPT-----DSISQLEASTSTSTYTRVA- 100
DB 341 VAFPRDSKLTFT--LKDS--LGNSKTFMTATVSPSALNTEETLSTLRVSRARDIYNVAQ 397
QY 101 -----AKDYDEAKSNFTAKSGLENA-----KTLAEYETKADIMAAIADOMERLA 145
DB 398 VNEDPRARRIRELEQEMEDMQMAGDPAYVSELKKLALLESEAKRAADIALER-- 455
QY 146 NSDPSNNHTEEVNNIKKALEAKOTIDKLNKLVTLQONOKSLTEVLTQSDAQIPAINS 205
DB 456 ---EREHNOVERLLRATRAEAKSELE--SRAALOE-----EMTATRQADKQMALNI 503
QY 206 QT-EIKNSADQIIKDLERONISYEAVLTNAGEVIKASSEAGIKIGQALOSIYDADQSO 264
DB 504 RLKEBQARKERELKMAKDAALSKVRRKDAEIASERE--KLESTVAOLEREQERE 560
QY 265 AAV--LQAOON-----NSPDNIATKELIDAEETVNVNLEKQEHGTLTSPLYKKAEEQI 316
DB 561 VALDALQTHQRKLOALESSERTAAER--DOLLQOLTELQSERIQL--SQVVTDRRE-- 612
QY 317 SQAKNDIQIK-----PGSDIPIYGPSSGASASAGALKSSNNSGRISILLDDVDN 370
DB 613 -RTRRLQRIQYEGETELARVDALCAQOEMERYHAAVFILOT-----LLELATWED- 665
QY 371 ENMAIALQCFRSMIEQFNVPNNPATAKEILOAMEA-----OLTAMSDOLVGAD-- 416
DB 666 -----ALRE-RALAER-----DEAAAELDAASTSONARESCERLTLSLEQOLRESEERA 715
QY 417 GELPAETQAIKDA-----LAQALKQPSADGLATAMGOVAPPAKAVGG--GS 460
DB 716 AELASQLEETAAKSAEODRENTATLEQOLRESEARAELASQLEATAPAAKMSAEODR 775
QY 461 AGAGTGVQNNVQLYKTA--FSSTSSSYAAALS--DQSAVKTLNLSLYSESR----- 509
DB 776 ENTRATLEQDRSEARAELASQLESTTAAKASABODESTRATLEQOLRDESEERAEL 835
QY 510 -SGVQSAISQIOTANPALSRVSRSIGESQGRSADASORAAETIYRDSQTLGDVYSRIQVLD 568
DB 836 ASQLESTTAAKSAEODRENTATLEQOLR--ESERAAE-----LAS 876
QY 569 SLMSTTVSNPQANOE-----IMQKLTASISKAPQGYPAVONSADSLQKFAQLERE 622
DB 877 QLESTTAAKSAEODRENTATLEQOLRDESEERAEL--ASQLEATAPAAKSAEODREN 933
QY 623 VDG--ERSLAEQENA 636
DB 934 TRAALEQOLRDESEERA 949

RESULT 22
US-08-282-845-2
Sequence 2, Application US/08282845
Patent No. 5719263
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: A 230kD Antigen Present in Leishmania
TITLE OF INVENTION: Species
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Macintosh Operating System 7.1
SOFTWARE: Microsoft Word for Macintosh 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/282,845

FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/006,676
FILING DATE: JANUARY 15, 1993

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 5004-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 955 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-282-845-2

Query Match 5.7%; Score 182.5; DB 1; Length 955;
Best Local Similarity 21.5%; Pred. No. 0.00019;
Matches 158; Conservative 122; Mismatches 269; Indels 187; Gaps 36;

QY 10 SGPESVSSQSSMNPINQIANSSEKSTK-----ASEASPSASSVS 54
DB 292 AGSERVAQSQ-----VEGQ-----QFKETHINISLTLGRVIDYLDAMATKGAQAQYS 340
QY 55 SMSFLSSAKNALISLDAIILNKNS-----SPT-----DSISQLEASTSTSTYTRVA- 100
DB 341 VAFPRDSKLTFT--LKDS--LGNSKTFMTATVSPSALNTEETLSTLRVSRARDIYNVAQ 397
QY 101 -----AKDYDEAKSNFTAKSGLENA-----KTLAEYETKADIMAAIADOMERLA 145
DB 398 VNEDPRARRIRELEQEMEDMQMAGDPAYVSELKKLALLESEAKRAADIALER-- 455
QY 146 NSDPSNNHTEEVNNIKKALEAKOTIDKLNKLVTLQONOKSLTEVLTQSDAQIPAINS 205
DB 456 ---EREHNOVERLLRATRAEAKSELE--SRAALOE-----EMTATRQADKQMALNI 503
QY 206 QT-EIKNSADQIIKDLERONISYEAVLTNAGEVIKASSEAGIKIGQALOSIYDADQSO 264
DB 504 RLKEBQARKERELKMAKDAALSKVRRKDAEIASERE--KLESTVAOLEREQERE 560
QY 265 AAV--LQAOON-----NSPDNIATKELIDAEETVNVNLEKQEHGTLTSPLYKKAEEQI 316
DB 561 VALDALQTHQRKLOALESSERTAAER--DOLLQOLTELQSERIQL--SQVVTDRRE-- 612

```

RESULT 23
PCT-US94-00324-1
: Sequence 1, Application PC/TUS9400324
GENERAL INFORMATION:
: APPLICANT: Reed, Steven
: TITLE OF INVENTION: Diagnosis of leishmaniasis
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Immunex Corporation
: STREET: 51 University Street
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98101
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Apple System 7.1
: SOFTWARE: Microsoft Word, version 5.1a
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/00324
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/006,676
: FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
: NAME: Perkins, Patricia Anne
: REGISTRATION NUMBER: 34,693
: REFERENCE/DOCKET NUMBER: 5004-WO
TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 587-0430
: TELEFAX: (206) 233-0644
: INFORMATION FOR SPO. ID NO.: 1
: SEQUENCE CHARACTERISTICS:
: LENGTH: 955 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-00324-1
Query Match 5.78; Score 182.5; DB 5; Length 955;
Best Local Similarity 21.58; Pred. No. 0.00019;

```

```

Matches 158: Conservative 122; Mismatches 269; Indels 187; Gaps 36;

QY 10 GCPESVSSNOSMPITNGQIANSSETRK-----ASESPASSSVS 54
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 292 AGSERVAVSO-----VSGO-----QFKEATHINLSLTTLGRVIDVLADMTAKKAQYS 340
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 55 SMSFLSSKKNLISLRDLINKNS-----SPT-----DSLQLEASTSTSVITRA- 100
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 341 VAPFROSKLFTI--LKDS-LGGSNKTFTIAFVSPALNYEETLSTLRYSARDDIVNAQ 397
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 101 -----AKDYEAKSNPDTAKSGLENA-----KTLAEYTKMADLMALMERIA 145
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 398 VNEDPRARIRIELEQMDMQAMAGBPAYVSELKTKLALLESPAOKRADLQALER-- 455
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 146 NSDSPSNHTEEVNNIKKALEAOKDTIDKLNLVTLQONKSLTEYLKTTSDAQIPATNS 205
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 456 -----EREHNOVOERLIRATEEKSLEL--SRAALQE-----EWTATRQADKKQALNLT 503
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 206 QL-EINKNSAQIITKDLERONISYEAVLTNAGEVYIKASSEAGITLQALOSIVDAQSO 264
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 504 RLKEQOAKREBELLEKMAKKDAALSKVARRRDAELASERE--KLESTVAOLREBQERE 560
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 265 AAV--LQAQN-----NSPDNIAATKELIDPAETKVVELKQEHHTGLDSDPLVKAAEQI 316
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 561 VALDALQTHQKLOEALLESSERTAER--DQLLOQLTELQSERTQL--SQVYIDRE--- 612
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 317 SOAQKDIOEIK-----PSGSDIPIVPGSGSASGSAAGALKSSNNGRISLLLDVYN 370
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 613 -RLRFDLRQIYEGEFTELARDVALCAQEMEARVHAHVFIQT-----LLELATWED- 665
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 371 EMAIALLOGFSSMLEQFVNNPATAKELQAMEA-----QLTAMSQLYGAD--- 416
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 666 -----ALRE-KALER--DEAAAALDAAASTSQNAREGACERLTLSEQLRESEERA 715
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 417 GELPAIDAIKDA-----LQALQOPSDAGLTAMGOVAFAAKVGCG--GS 460
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 716 AELASQLEATYAAKSSAEOBRENTRATLEQQLRESEARAALASQLEATYAAKSSAEOBR 775
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 461 AGTAGTVQMNVKOLKYTKA--PSTSSSSSYAALSS--DGSAVYTLNLSYSESR----- 509
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 776 ENTRATITLQQLRDSERRAELASQLESTYAAKSSAEOBRENTRATLEQQLRDSERRAEL 835
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 510 -SGVQSAISQTPANPALRSVSRSGIESQGRSADASQRAAETTVRDSQTLGDVYSRLOYLD 568
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 836 ASQLESTYAAKSSAEOBRENTRATLEQQLR--ESEERRAE-----LAS 876
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 569 SLMSIYSNPQANOE-----IMOKLTASISKAPQCFYPAVQNSADSLQKFAANQLEREF 622
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 877 QLESTYAAKSSAEOBRENTRATLEQQLRDSERRAEL--ASQLEATYAAKSSAEOBRENT 933
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 623 VDG--ERSLASQENA 636
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 934 TRAALEQQLRDSERRA 949
   | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 24
US-09-336-447A-5
: Sequence 5, Application US/09336447A
: Patent No. 6310190
: GENERAL INFORMATION:
: APPLICANT: HANSEN, ERIC J.
: APPLICANT: AEBI, CHRISTOPH
: APPLICANT: COBE, LESLIE D.
: APPLICANT: MACIVER, ISOBEL
: APPLICANT: FISKE, MICHAEL J.
: APPLICANT: FRDENBURG, ROSS A.
: TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
: FILE REFERENCE: AMCY.024
: CURRENT APPLICATION NUMBER: US/09/336, 447A
: CURRENT FILING DATE: 1999-06-21
: NUMBER OF SEQ ID NOS: 98
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 5

```

LENGTH: 892
 TYPE: PRT
 ORGANISM: Moraxella catarrhalis
 US-09-336-447A-5

Query Match 5.7%: Score 182; DB 4; Length 892;
 Best Local Similarity 19.3%: Pred. No. 0.00019;
 Matches 126; Conservative 110; Mismatches 240; Indels 178; Gaps 28;

QY 7 HMEGPEVSVSSNOSSMNPIT-----NGOIASNSEFK-----STRKASESPSASS 51
 DB 271 HMTAGKATAVASAKVNGLTILGNFAGASKTGNGTVSGSENNRQIVNAGNISADSTD 330
 QY 52 SVS-----SMSFLSSAKNALI-----SLDALINKSSPTDISLQLEAST 91
 DB 331 AVNGSOLVALATAVKADADENFALTKTONTLEGEADADALQON-----QTDITPA 382
 QY 92 STSTVTRVAKDYDEAKSFND-TAKSGLENAKTLAEYETRMADLMALQDM-ERLAN--- 146
 DB 383 NKTAIR-----NFNRTVVGFELEKKNKAGIAKNQADITQTLNNGEELLNLISG 431
 QY 147 -----SDPSNN-----HTEEVNNIKALEAKDTIDKLNKLVTLQONKSL 187
 DB 432 RLDDQKADIDNNINITYDLAQODQSSDIKTKKNVE--EGLLDLSGRLL--DQKADL 486
 QY 188 TEVLKT-----TDSADQIPAINSOLEINKNSD-----QIYDLERONISTYAVLT 233
 DB 487 TKDITKLENNVEGLDLSGRLL-IDQKADIAKNQADIAONQTDIDDLAAYNELQDYAQ 544
 QY 234 NAGEV-----KASSEAGIKGALQALQISYVAGDQSOAAVLOAOONNSPDNIATKELIDAA 289
 DB 545 KQTEAIDALNKASSANTDRIATATLGT--AENKKQAIKAKQAKENKKDGIKAKQADITQLH 602
 QY 290 ETKVNELKQEHGTL-----TDSPLYKAEQISOAKDIOELKPS---GSDI 333
 DB 603 DKKITNLGILHSVAVRAGVNGNTQGVATNKADIAKNQADIANINKIYELAQODQSSDI 662
 QY 334 PIYPGSGSAASAGSAGALKSSNNGRISILLDDVDNEMAALQGFPSKITEFPANNNA 393
 DB 663 -----KTLAKVSAANTDRIAKNKAADASFETLTKRNOTLTLEQG-EALVEQ---NNA 710
 QY 394 TAKELOAMEAQLTAMSDOLVAGDELPAETIOAIKDALAQALQOPASGLATTAMGOVAFAA 453
 DB 711 INDELGFPAHADVDQKOLL-----QNOADITTKKAAIEQNIINTVANGF----- 755
 QY 454 AKVGGSGAGTAVQW-----NVKQLYKT--AFSSSTSSSYAALSDGYSAYKTINSL 504
 DB 756 -EIEKKAGIATKQELLIONDRLOINETNNRQDKIDOLGYALK-EGQGHNNRISAV 813
 QY 505 YSRSRGVOSAIQOTANPALSR-----SYRSRGIESQGRS 539
 DB 814 ERQTAGGIANALAIATLPSPSRAGEHHVLFGSGYHNGQAVALSAGAGLSDPTGS 867

RESULT 25
 US-08-328-254-6
 ; Sequence 6, Application US/08328254
 ; Patent No. 5710022
 ; GENERAL INFORMATION:
 ; APPLICANT: Zhu, Xueliang
 ; APPLICANT: Lee, Wen-Hwa
 ; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell and Flores
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/328, 254
 FILING DATE: 24-OCT-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/141,239
 FILING DATE: 22-OCT-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-CJ 1191
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2482 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-328-254-6

Query Match 5.6%: Score 180; DB 1; Length 2482;
 Best Local Similarity 19.9%: Pred. No. 0.0011;
 Matches 151; Conservative 125; Mismatches 288; Indels 196; Gaps 32;

QY 13 ESVSSNOSSMNPILINGQISNSEFKESTKASEASPSASSSVSSWFLSSAKNALISLRDA 72
 DB 317 ETLSELEKKESSITS---LNKREIELTQENGTLKRNKS-----LNQKNNLLOKSS 367
 QY 73 LKNNSSPTDLSQ-----LEASTSTVTRVAAKDY-----DE 106
 DB 368 FANYIDERKKSISELSDQYKOEKLLQRCFEFGNAYEDLSQYKKAQKNSKLECLINE 427
 QY 107 AKSNEDTAKSGLEN-----AKTLAEYETKMA-----DLMALQDMERLANS---DPSN 151
 DB 428 CTSLCENRRNEQLKEAPRAKEHOFELTKLAFABERNQNMLELETVQDALRSEMTONON 487
 QY 152 NHTEEVNNIKALEAKQKDTIDKLNKLVTLQONKSLTFLVLTQTSADQI---PAINS-- 205
 DB 488 NKSSEAGIGKQELMTLKEQONKQKQEVNDLQENQELKVMKTKHQCQMLESPFINSVK 547
 QY 206 -----OLEINKNSADQIYDLERONISTYAVLTNAGEVITKASSEAGIKG 250
 DB 548 ERESERNQCNFKPQMDLEVEKEISLD---SYNQLVQLEAMLRNKLKLDQESKEKECLO 603
 QY 251 QALOSIVDAGDQSOAAVLOAQONNSPDNIATKEL-IDAAE----- 290
 DB 604 HELQTLRGDLETSNLODMQSOE-----ISGLKDCETDAEIKISGFHELSTSONDAHL 657
 QY 291 -----TKVNELK-----QEHGTGLTDS-----PLVKAQEOISOAKDIOET 326
 DB 658 QCSLQTTMKNKINLEKICELQAEKYLEVTLELNDSEECITATRKAAEEVG---KLINLV 714
 QY 327 KPSGSDIPIVG-----PSGSAASAGS-----AAGALKSSNNGSITISILLDDVDNEMA 374
 DB 715 KIINDSGLLHGLVLDIPGGEFGEPQNEQHPVSLAPRLDESSNSYEHLLTLDREVQNHFAE 774
 QY 375 -----IALQGFRRMIEQOPVNNPATAKELQ---AMEAQLTAMSDOLVAGDELPAEIO 424
 DB 775 LQKFLSLQSEHKTLHDQHCQSSKSEIQTYVDSLKAELVLYSTLRNFGQGLVAKEMOL 834
 QY 425 AIKDALAQALQK---PSADGLATAMGOVAFAA---KVGGS-----AGTACTVQNNVNO 473
 DB 835 GLEEGLVPSLSSSCVDPSSSL--SLQDSSFRYRLLEQTDQMSLLSLDEGAVSANQCSYDE 893
 QY 474 LYKTAFF--SSTSSSYAALSDGYSAYKTINSLYSRSRGVOSAIQOTANPALSRVSNS 531
 DB 894 VFCSSLOEENLTKRTPAPAKGVLEELISCEVYRQSLKLEK--- 937
 QY 532 GIESQGRSADASORAETIVYRSQTLGDVYSRLQVLDSLMTSTIVSNPQANQETIMQKTLA 591

Db 938 -MESGIMKNEIOLLEOLLS-----SERQELDLRKOYLS-----ENFOWOOKLTS 983
Oy 592 -SISAKPOFGYPAYONSADSLOKFAOLEREVD-GERSL 629
Db 984 VLEMEKSLAAEKKOTEOLESLEVARLOGLDLSRSL 1023

RESULT 26

US-08-685-871-2
Sequence 2, Application us/08685871
Patent No. 6013499
GENERAL INFORMATION:
APPLICANT: NARUMITA, Shuh
APPLICANT: IWAMATSU, Akihiko
TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685, 871
FILING DATE: 24-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184102
FILING DATE: 25-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-262553
FILING DATE: 14-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16887/845
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-685-871-2

Query Match 5.6%; Score 178.5; DB 3; Length 1354;
Best Local Similarity 19.8%; Pred. No. 0.0006; Indels 179; Gaps 34;
Matches 141; Conservative 123; Mismatches 269;

Oy 12 PESVSNOSNNPIINGOIASNSEFEKSTKASEASPSASSVSSWSFSSAKNALISLRD 71
Db 384 PKAVYGNL---PEYGFYYNRRRLSSANND---NRTSSNADKSLQESLOKTYKLEE 437
Oy 72 AILKNSSPDTLSOLEASTSTSTVTRAAKOYDE---AKSNFTAKSGLENAKTALEY- 127
Db 438 QLAHNE-MOLKDEMRO-KCRTSNIKIKDKI-MKELDEGQNRRLLESTVSOIEKEMKLLQHR 494
Oy 128 -----ETKADIMAAALODMERLA-NSDPSNNHTEEVNNIKKALEAQK 168
Db 495 INEQRAKAEQENKRRNVEVSTLKQLEDLKKVSONQLAN---EKLSQLQLEBAN 551
Oy 169 DTIDKLNLVTLQONKNSLFEVLKTTDSADQIPAINSOLE---INKSADQIITKD----- 220

Db 552 DUL--RTESDPAVRLRKSHTEMSK---SISOLESLNRELQERNRLILENSKSQTDKDYVL 606
Oy 221 -----LERONISY-----EAVLTNGEVIKASSEAGITLGOLOSIVAGOSOAAY 267
Db 607 QALIEARRDRKHDSSEMIGDILQARITSLQEEVK-----HLKHNLEKV--EGEKREKAD 657
Oy 268 LOAOONNSPDNIA-----ATRELIDAEFKYNELKOEHTGLTD-----S 306
Db 658 MLNHSKKNNLEIDLANKKLKSLQGRLEQEVNEHKVTARLTDKQSTIEAKSVAMCEME 717
Oy 307 PLVKAEPQISOAKODIOEIKPSGSDPIYGPSSASAGSAGALKSSNN-----SGRI 361
Db 718 KKLKEERBARREKAEENRVVQIKQCSMLDY-----DLKOSQOKLEHLTGK 762
Oy 362 SLILDVQNEAAIATALOFRSMIRDFNNPATAKEL--QAMENQLTMSQLVADDEL 419
Db 763 ERMEDEVKN---LTLQ-----LQESKKRLLONELKTOAFEA---DNLGKLEKQ 807
Oy 420 PAEIOAIKDA-----LAQALKOPADGLATAMGOVAFAAKAVGSGSAGTAGTV----- 467
Db 808 KOEINTLLEAKRLLEFELAQLTKO-----YRGNEGQKRELD 844
Oy 468 QMNVKOLYKTAFTSSTSSSYAALSDGYSAYKTLNLSYSESRS-GVQSAISOTANPALSR 526
Db 845 QLEAEQYFSTLYKTOVKELKEIEEKKRNENLKIOELQNEKETLATQDLAET--KAESE 902
Oy 527 SVSRGITSQGRSADASGRAETIYRDSOTLGD---VYSRIQVDSLMTSTIVSNQAOE 583
Db 903 QLARGLLEEQ--YEELQESKKAASRNHOETTDHVTSLSEANSMTKIEILRRENE 960
Oy 584 EIMOKLTASISKAPQFGYPAYONSADSLOKFAOLEREVDGERSLSAESQEN 635
Db 961 ELTEKM-----KKAEETKL--EKEEISNLKAEKFN-INTERTLKQAVN 1004

RESULT 27

US-09-310-187A-1
Sequence 1, Application US/09310187A
Patent No. 6358751
GENERAL INFORMATION:
APPLICANT: Benichou, Gilles
APPLICANT: Fedoseyeva, Eugenia
TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
FILE REFERENCE: UCSF-090
CURRENT APPLICATION NUMBER: US/09/310,187A
CURRENT FILING DATE: 1999-05-12
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1939
TYPE: PRT
ORGANISM: Homo sapiens
US-09-310-187A-1

Query Match 5.5%; Score 175; DB 4; Length 1939;
Best Local Similarity 19.7%; Pred. No. 0.0017; Indels 142; Gaps 28;
Matches 130; Conservative 125; Mismatches 262;

Oy 35 ETKESTKASEASPSASSVSSW-----SFLSSAKNALIS-LRD-----ATLN 75
Db 1352 ETEKAEILQVRLSKANSYVAQMRKYETDAIORTBELEBAKKLQARLODAEVAEAVNA 1411
Oy 76 KNSSPDTLSOLEASTSTSTV---TRVAAKOYDEAKSNFTAKSGLENAKTALEYETKM 131
Db 1412 KCSSELEKTKHRLQNEIEDIADWVERSNAAALDKKQHNFD-----KILEMKQKY 1462
Oy 132 ADLMAALODMERLANSDPSNNHTEEVNNIKKALEAQKOTIDKLKIYTLQONKNSLFEVL 191
Db 1463 EESOSELESSQKEARS-----LSTELFKLNAVE-----ESLHLETFKRENNLOE-- 1509
Oy 192 KTTDSADQIPAINSQL-EINKN--SADQIIDLFRONISYEAIVLNA-----GEVIK 240

```

Db 1510 -----EISDLTEQLGEGKNVHELEKVRKOLEVEKLELQSLAEBAEASLEHEEGKIILR 1562
QY 241 ASSEAGIKLGOALQSLQIVDAGDOSQAALVLOAOONNS--PNNIATKLELIDAEETKVELKO 298
Db 1563 AOLEFNOIKAETIERKLAEDDEME---QAKRHQRYVDSLOTSLDAETRSRNEVLRVKK 1618
QY 299 EHTG-LTDSPL-VKKAEOISQAQKDIOEIKPSGSDIPVPGSGSAASAGSAAGLKSSN 356
Db 1619 KKEGDLNEHEIDQSHANRMAEAQKQVKSLSGLKDTQI-----QJDDAVR-AN 1666
QY 357 NSGRISLLDDVDNEMALALALOGFRSMIEQFVNNPATKEL-----QAMBOULTAMS 409
Db 1667 DDLKENIAIVERNNLLOAEELRAVVBOTERSKRLAOLETSEVQLHSQNTSLI 1726
QY 410 DQ-----LYGADGELPAEIDAIKDALQALQKPSADGLATPAMGVAPAAKVGSGAG 462
Db 1727 NOKKKNESLUTQLOSEVEEAQVCECRNAEKKAK-----AITDAAMAEELKEQDT 1777
QY 463 TAGTQVM--NVKQLYKTAESSTSSSYAALSDGYSAYKTLNLSYSESRSGVQSAISQTA 520
Db 1778 SAHLERKKKMEQTIKD-LQHRLEDAEQIALKGKKQLQKLEARVRELGELEAEQKRMA 1836
QY 521 NPALSNSVRSGL-ESQGSADASQRAETIVRDSOTLGDVYSRLQVLDLSMSTIVSNQ 579
Db 1837 E-----SVKGRKSERIKELTYOTE---DKRNL-----LRLODLVKLOLKYKAYK 1881
QY 580 ANOEIMOKLTASISKAPQGYPAVONSADSLQKFAQLEREFVDEGRSLAESQENAFR 638
Db 1882 RQAEFEAEQANTNLSK-----FRKVOHELDEAEERA-----DIAEQVAKLR 1923

```

RESULT 28

```

US-09-268-347-36
; Sequence 36, Application US/09268347
; Patent No. 6353182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2411
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-36

```

```

Query Match          5.4%; Score 174; DB 4; Length 2411;
Best Local Similarity 19.3%; Pred. No. 0.0028;
Matches 143; Conservative 123; Mismatches 307; Indels 166; Gaps 30;

```

```

QY 1 MHHHHHMSGPESVSSNOSSMNPINGOIAS-----NSETKESTRASEA 45
Db 562 LNRHVEDAYKGLNLEKNKANKQPLVDSYAATVGLRKLGVVSTKNGTKESNQVKA 621
QY 46 SPBASSSVSMFLSSAKNALISL-----RDALINKNSPTDSLQLEASTSTSTY 97
Db 622 DEVLFTGAGAAVTYKSKSENGKHITVSAETKADSGLEKDG---DTIKLKVDMQNDNVL 678
QY 98 RYAAKDYDEAKSNFTPAKSGLENA-----KTLAEYETKMA---DLMAALOD----- 140
Db 679 TYGNNGTAVTKGGFETVTKGATDADRGKYTVKATATANDADKKAAVYKDVATATNSATVY 738
QY 141 -MERLANSPPSNHTEEVNNIKKALEAKDQTDIDKLNKLVTLQONKSLTEVL-----KT 193
Db 739 KTEMLFTSIDENPTD--NGKDALAKAGDPLFPKAKKNLKVKRDGNITFDLAKNIEVK 796
QY 194 TDSADQI-----PAIN-----SOLEINKNSADQIIKDLERONISYEAVLN 234
Db 797 AKVSDTLITIGNTPTGTATPKVNITSTADGINFAKETADA-----SGSKNVYLKGIATY 852

```

```

QY 235 AGEVIRKASSEAGIKLGOALQSLQIVDAGDOSQAAL-----VLOA-----QONNSPDNIATKELI 286
Db 853 LREPSAGAKSHVLDL-----VDATKKSNMAASIEDVILRAGNNIDQNGNNVYVATYDTV 906
QY 287 DPAETKVELNKOEHGTLTSPLYKKAEOISQAQ-----KDIOET--- 326
Db 907 NPTD-----DSGTFTTVYQKADGKADGVKIGAKTSVIKDHNGKLETKDKLDANN 958
QY 327 -----KPSGSD-----IPIVGPS-----GSAASGASAGALKSSNSGRISLL 365
Db 959 GATVSEDDCKDTGTGLVTAKYTVIDAVNKSQKRYTGEGATAEFGATRAVNAAGMETYTSGIS 1018
QY 366 DDVDNEMALAL-----QGRSMIEQFVNNPATKELQAMEQOLTAMSQVLGADGELPA 421
Db 1019 VNFKNKNATPATAVSKDNGNINVKYDVNVGDGLKIGDKKIVADTTTLT--VYGGKVSVA 1076
QY 422 EIQAIDALQALQKPSADGLATAMQVAFPAAKVGGSGAGTAVQMMNVKQLYKTAES 481
Db 1077 GANSVNN---NKKLVNAEGLATALNNLSMTAKADRYADGESEGETDQEVKAGDKVTFKA 1132
QY 482 TSSSSYAALSD-GYSAYKTLNLSYSESRSGVQSAISQTAANPALSRVSRGIE---SOG 537
Db 1133 GKNLKYKQSEKDEFTYSLQDTLGLTISITLGTANGRNDTGT-----VINKDGLITILANG 1187
QY 538 RSA--DASQRAETIVRDSOTLG---DYVSRLQVLDLSMSTI-VSNPOANOEIMOKL 589
Db 1188 AAGGTASNGNTISVTKDGISAGNKEITVVKSKLTKYKDTQVTAGATQPAAMTAEVAKOD 1247
QY 590 TASISKAPQGYPAVONSADS 610
Db 1248 LVDLTK-PATG--MAGNGADA 1265

```

RESULT 29

```

US-08-938-105-3
; Sequence 3, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Leinwand, Leslie A.
; APPLICANT: Vikstrom, Karen L.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,105
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crook, Wanneil M.
; REGISTRATION NUMBER: 31,071
; REFERENCE/DOCKET NUMBER: 3595-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1886 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-938-105-3

```

Query Match	5.4%;	Score 173.5;	DB 4;	Length 1886;
Best Local Similarity	21.7%;	Pred. NO. 0.0021;		
Matches 133; Conservative	86;	Mismatches 236;	Indels 159;	Gaps 26;

Qy	86	QLEASTSTVTVRAAKVDYEAKSNEFDIAKSGLENAKTLAEYETMMADMAALODMERLA	145
Db	863	QLEAKVKEW---ERLEDEEEMAEJLTKARRKLEDECSLKKDIDLEJTLAKVEKEHA	919
Qy	146	NSDSNNHTEEVNNI-----KKAL-EAOKDITDL---NKLVTLONQKSLREY	190
Db	920	TENKVKVLTTEAMGLDEITLAKLTKEKALQEAHQALDLOAEERKVTITLTKSKYKLEO-	978
Qy	191	LKTTDSADQIPAINSOLEITKNSADQILKDERONISYEAVLTNAGEVYIKASSEAGIKLG	250
Db	979	-----QYDDEGSLSEGEK---KYRMDLERAKKLEGLDKLQESIMDLENDKIDLE	1026
Qy	251	QAL-OSIVDAGDS-----QAAVLOAQOONSNDPNTAAKELIDAEFKVMELOEHTGL	303
Db	1027	EKLKKKEFDISQONSKITEDBOLALOLOL-KUKENQARILEE-----ELLEKRTAR	1078
Qy	304	TDSPLVKAEEQISOAKODIOELIKPSGSDPIVGPESGAASGAALGKSSNNGRISL	363
Db	1079	A-----KVEKLKSDLTJRELEI-----SERLEEGGA-----TSYQJEM	1112
Qy	364	LDDVDNEMAAIILGCFRSTEDOFVNNNPATKELQAMEAQLTAMSDDLVGAGELPAEI	423
Db	1113	-----NKKREAFQKRRRLDEATLTQHEATLTAAALRKKHADSV---EL---GEQIDNL	1159
Qy	424	QAIKDALQALQLO--PSADCLATAMGOVAFPAAKVGGSGAGTGVONNKOVLTKTAPSS	481
Db	1160	QWVKQKREKESKFEKLELDVYSHMQIITKAA-----NLEKVSRTL--E	1202
Qy	482	TSSSSYAAALSDGYSAYKTLNSLYSESRSQVSAISOPTANPALSRSSVS-----	529
Db	1203	DOANEYKVKLEE---AORSIMFTT-----ORAKLTQENGELARQLEEKBALLIQLTRG	1253
Qy	530	-----RSGISQSGRSDASQRAAEFTVPSQFTLGVDYVSKYLOVLSIMSTVTS--	576
Db	1254	KLSTQOMEDLKQOLEEBGKKAALNHALQASARHDDLLIREQYEEEMERAKKELORVUSKA	1313
Qy	577	NPOANO-----EELMOKLTASISKAPQFGYRAVONASADSLOKFAAQRD	620
Db	1314	NSEVAGQRTYETFDIAQRTQEELLEEAKKTLAQRLODAEE-AVEAVNAKSSLEKTKHRLQN	1372
Qy	621	E-----PVDGERSLA	630
Db	1373	EIEDLMAVDERSNA	1386

```

RESULT 30
US-09-336-447A-13
? Sequence 13, Application US/09336447A
? Patent No. 6310190
? GENERAL INFORMATION:
? APPLICANT: HANSEN, ERIC J.
? APPLICANT: AEBI, CHRISTOPH
? APPLICANT: COPE, LESLIE D.
? APPLICANT: MACIVER, ISOBEL
? APPLICANT: FLSKE, MICHAEL J.
? APPLICANT: FREDENBURG, ROSS A.
? TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
? FILE REFERENCE: AMCY:024
? CURRENT APPLICATION NUMBER: US/09/336,447A
? CURRENT FILING DATE: 1999-06-21
? NUMBER OF SEQ ID NOS: 98
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 13
? LENGTH: 873
? TYPE: PRN
? ORGANISM: Moraxella catarrhalis
? US-09-336-447A-13

```

Query Match 5.48; Score 172.5; DB 4; Length 873;

Best Local Similarity 20.78; Pred. NO. 0.00083;
Matches 124; Conservative 97; Mismatches 238; Indels 139; Gaps 27;

[illegible]

```

1 RESULT 31
2 US-09-103-664A-2
3 : Sequence 2, Application US/09103664A
4 : Patent No. 6458358
5 : GENERAL INFORMATION:
6 :
7 : APPLICANT: University of Kentucky Research Foundation
8 :
9 : APPLICANT: Timoney, John
10 :
11 : APPLICANT: Atliushin, Sergey
12 :
13 : TITLE OF INVENTION: Compounds Encoding the Protective M-Like Protein of Streptococ
14 :
15 : TITLE OF INVENTION: and Assays Therefor
16 :
17 : FILE REFERENCE: 50229-212
18 :
19 : CURRENT APPLICATION NUMBER: US/09/103,664A
20 :
21 : CURRENT FILING DATE: 1998-06-23
22 :
23 : NUMBER OF SEQ ID NOS: 10
24 :
25 : SOFTWARE: PatentIn version 3.0
26 :
27 : SEQ ID NO 2
28 :
29 : LENGTH: 534
30 :
31 : TYPE: PRT
32 :
33 : ORGANISM: Streptococcus equi
34 :
35 : US-09-103-664A-2

```

Query Match	5.48;	Score 172;	DB 4;	length 534;
Best Local Similarity	22.0%;	Pred. No. 0.00045;		
Matches	94;	Conservative	80;	Mismatches 166;
			Indels	88;
			Gaps	16;

```
Qy      69 LRDAILNNSSPTDLSLQLEASTST--SYTVRAAKDDDEAKSNFTAKSLEENATLAELAK 126
        ||: | : : : ||: ||: | : : : ||:
Db     169 LRNLKEKEQEQLTRAKDERGALTSEFNKTLIR-STKEVKKLKTE-----LAK 214
```

[illegible]

RESULT 32
US-09-336-447A-15
; Sequence 15, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PR1
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-15

Query Match 5.3%; Score 171; DB 4; Length 889;
Best Local Similarity 21.6%; Pred. No. 0.0011;
Matches 130; Conservative 82; Mismatches 226; Indels 164; Gaps 24

OY	14	SVSSXOSSMNFILINGQILANSSETE-----	STRKASEPSPAS--	SSVSXMSF	58
Db	359	TVKKKOK--NFILG---SNTDITKDOAGSVLLDNTMGRKATVVEDYDGLSLTFPAG		4133	
OY	59	LSAKNALISL-----RDAI-LKNSSPYTSLSOLEASTSTSVTRPAADKYDEAKSNF		1111	
Db	414	VSKANSIGSVYSGSEKKEHQIVHVGGRISNSTDVAVNGSOLYALAAVDDNQYDIEKNOD		4734	
OY	112	DTAKSGLENAKTLAEETFKMADLMAALODMERLANSDBSNHTEEVNINKALEAQODTI		1717	
Db	474	DIAKNOADIANKQADIQTL-----ENDVGKELLNLSGRLLDOKADI		5141	
OY	172	DK-----LKLVTLOQNRSLTEVLTCTDSDAQIPAINSOLEIINKNSDQ---IIRKDE--		2222	
Db	515	DNNHITIELLQAOODDQHSDDIKTLKKN-----VEBIELELSGHLIDOKKADILKIDIKAL		5666	
OY	223	KONISYEAVLITNAGEVIR-----ASSEAGIKLGALOSIYDAGDOSQAALVLOAQONNSPD		2717	

Db	568	ESNVE-EGLLDLSGRLLDQKADIAQNOQANIQDLAAAYNELDQYAAQKQTEAIDALNKASSE	626
Qy	278	N-----IAAKELIDA-----AEKRVNELQEHGILTDSPLVYKKAEEQISOAQKDI-	323
Db	627	NTONIEDLAAVNELDQYAAQKQTEAIDALNKASSENT-----ONIAKNQADIA	674
Qy	324	-----OEIKPSCSDIPIYPPSGSAAGSAGAAALCKSSNNSGRISILLDDVQENEMA	373
Db	675	NNNNIYELAQDQDQSHSDIKTLAKASAA-----NTDIRIAKNKADAD-----	716
Qy	374	AIALQGRSMTEQTNVNNNPATAKELQAMEKQUTAMSQLYGADGELPAELQAI-----	426
Db	717	-----ASFETLLKNONTLEIKDEHDKLITANKTAIDANKASAPTKFATADATTKNGNAL	772
Qy	427	-----KDALAQALQKPSADGLTAM-----GOVAFPAAKVGGSGSAGTA---GTVO--MN	470
Db	773	TKNAKSTTDLQTKVDGEDGRYALTDLDTVNAFDGKITLALDSKVENGMAQAALSGLRQPS	832
Qy	471	VKQLYKTAFSSTSSSYAAALSDGYSAYKTLNLSLYSESRSGVOSAIISQTNANPALSRSVSR	530
Db	833	VGFENATPALGQSGSKSAVAIGAGYR-----VNPDLAFKGAAGAAINTSGNK	877
Qy	531	SG	532
Db	878	KG	879

```

: RESULT 33
: US-09-134-001C-4318
: Sequence 4318, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucelte-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIOR FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 4318
: LENGTH: 676
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-4318

```

Query Match	5.3%;	Score 170.5;	DB 4;	Length 676;
Best Local Similarity	18.2%;	Pred. No. 0.0008;		
Matches 108;	Conservative 111;	Mismatches 235;	Indels 139;	Gaps 18;

QY	106	EAKSNEDPAKSSNAKTLAEYETEMADLMALOMERLANSPPSNHHEEVNNIKKALE	165
Db	32	DAFPOSESTSSSWEVTSNSSNTADXIQQAKDKIDIKELISDD	IKSE 79
QY	166	AKQDTIDKLNKLVTLONOKSLTEVILKTTDSADQIPAINSOLEINKNSA	---DQIIKDE 222
Db	80	ERLDKVDNQSSIDRIITNDAPKRNHLLKSPDS	---ATSSKTEDDDTSEKNDMDMKID 135
QY	223	RONISEAVLTNAGEYIKRASSFAGIKLGALOSIYDAGGQSAAVLQAOQNNSPDIIAT	282
Db	136	KILSDIDSTAKVNDNQGENSA	---SKPDSYTDKEDKSNKNV---HDTNASTRNATT 188
QY	283	KELIDAEETKVELLQOEHHGIDTSPLYKKAEBQISOAQKDIOELKPSGSDIPIVGSQA	342
Db	189	DDSESVIDIKLDKIQDDFKSDSNMKLSQDSQDQASPKNENKNEES	-----ST 237
QY	343	ASAGSAGLAKSSNNSGRISLL	-----LDDVDNEMAI---ALOG--- 379
Db	238	TYNQSSDSKDKSDNGRSTLERIASDTPDJIRSKDQHYHTBEKODIQATISLQSGSKI	297


```

; APPLICATION NUMBER: US/08/125,077
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3111 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-125-077-4

```

Query Match 5.3%; Score 170.5; DB 2; Length 3111;

Best Local Similarity 19.9%; Pred. No. 0.0069; Mismatches 259; Indels 157; Gaps 26;

```

; Matches 139; Conservative 103; Mismatches 259; Indels 157; Gaps 26;

QY 13 ESVSSNOSNNPIINOIASNSETKSTKASEASPSASSSSVSWFLSSAKMALISLRDA 72
DB 1858 EDIOTKLPRSEELNKRIDLQSOEIKDRKLAERVSQASHA-----A 1899
QY 73 ILKNNSPPTLSOLEASTSTVTYRAADYDEAKSNFTAKSGLENAKTLAEYETKMA 132
DB 1900 QLDSSAVLDGI--LDEAKNISFNATPAFAAYSNIKDYIDEAEKVAEAKDLHAETKLA 1957
QY 133 -----DLMAALQ-----DMERLANSDESNHTEEVNNIKKALE----- 165
DB 1958 TGRGLIKEAKGCKLOKSPFLMEAKKLANDVENE--DHLNGLKTEIENADARNGDLR 2015
QY 166 AQRDTIDKLN-----KLVTLOQNKSLTEVLKTTDSADQIPAINSOLEINKNSADQI 217
DB 2016 TLNDFTLKLSAIFNDPAKLOAVKDKARQANDTAK--DVLQITELHQNDLGLKKYNKL 2073
QY 218 IKLERINIEEAVLTNAGE-VIKASSEAGIK-LGQALQSIIVDAGDSQAANVLOAQONNS 275
DB 2074 ADSVAKTN---AVVKDPSKNKTIADADATVKNLEQADRLID---KLKPIKELEDNL 2124
QY 276 PDNIAATKELIDAETKVNELKOEHTGLTDSPLVKKAEEQISOAQKDIOEIKPSGSDIP 335
DB 2125 KKNISEKELINARKQONIKVSSSGDCITRYKEIKKGSNNIYVNVKTRAVAD-NL 2183
QY 336 VGPSSAASAGSAGALAKSSNSGRTISLLD-----DVDNEMAAILAGFRS 382
DB 2184 LFLYLSAKFIDFLAIEMR---KGVSFLMDVSGVGRVEYPLDITIDS-----WYRI 2233
QY 383 MIOGFNNNATKELQAMEAOULTAMSDQLVGADG---ELPAEIOAKIKALQALQOPS 438
DB 2234 VASRTGNGTISYVALDGPFASTVPTNHTSPPGYTIILDVANAMLVFGGLGKILKAD 2293
QY 439 ADGLAT---AMGOVAFAAKVG-----GSGAG-----TACTVOMN----- 470
DB 2294 AVRYITFTGCMGETYFENKRIGLMNFREKGDCKGCTVSPQVESEBTIQFDGQYALVS 2353
QY 471 -----VKQYKTAESTSSSSSYAA-----LSDQSYAYKTLNLSYSESR 509
DB 2354 RPIRWYENISTVMPKFRFT-FSSSALLMYLATRLDRDFMSVELTFGHI-----KVSVDLG 2406
QY 510 SGVOSALS-OTANPALSRYSRSGIESQGRSA--DASQRAEFTIVRS 554
DB 2407 SGMAVSVNQNHNDGKWKSFLLSKRIQONISIVDIDTNOENIATSS 2454

```

```

RESULT 36
5352450-2
; Patent No. 5352450
; APPLICANT: KOGA, TOSHIHIKO; OKAHASHI, NOBUO; TAKAHASHI, ICHIRO;
; SHIBUYA, KOJI; OHTA, HIROAKA
; TITLE OF INVENTION: METHOD FOR PREPARING VACCINE FOR DENTAL
; CARIES AND VACCINAL COMPOSITIONS FOR DENTAL CARIES USED AS NASAL
; DROP
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/529,602
; FILING DATE: 29-MAY-1990
; SEQ ID NO: 2
; LENGTH: 1565
5352450-2

```

Query Match 5.3%; Score 169.5; DB 6; Length 1565;

Best Local Similarity 18.6%; Pred. No. 0.0031; Mismatches 267; Indels 117; Gaps 21;

```

; Matches 113; Conservative 111; Mismatches 267; Indels 117; Gaps 21;

QY 10 SCPEVSNNOSNNPIINOIASNSETKSTKASEASPSASSSSVSWFLSSAKMALISLR 69
DB 32 AGQKVFADETTTTSDVDKRVGTQGNPATQLPEAGSAS-----KEAQSOQTKL 82
QY 70 RDAIILKNNSPPTLSOLEASTSTVTYRAADYDEAKSNFTAKSGLENAKTLAEYET 129
DB 83 ERQWHTIEVPTKDLDQAAKDAKSAGYNNV--QDADYKCTVTPPEAVQKETEIEEDT 140
QY 130 KMAIDLMAALQDMERLANSDESN--NHTEEVNNIKKALEAQDTIDK-----LNKLVY 179
DB 141 KQA-----EDIKTTDQTKSDVAHAHEAEVAKIKAKNOATKEQERDMAKHAEVERINA 194
QY 180 LQONKNSLTFVLTQTTDSADQIPAINSOLEINKNSADQIILKLERONISIEAVLTNAGEV 239
DB 195 ANASTAFAEAKLAQYQAD-LAAVQ-----KTAANQO-----AAIQKALAAVQAE 239
QY 240 KASSEAGIKLQALQSIIVDAGDSQAANVLOAQONNSPDNIAATKELIDAETKVNELKOE 299
DB 240 KRYQENANMAAKAAYDPAVANAANNKNTIEAANBEIKRNTAKAEV---ETLQAYQAE 295
QY 300 HTGLTDSPLVKKAEEQ--ISQAKDIOEIKPSGSDIPYVGPSSAASAGSAGALAKSSNN 357
DB 296 LKRYQENANNEADYQAKTLAYQTELARQKANAD-----AKATYEAAVAANNAKN 346
QY 358 SGRTISLLDVIDNEMAILAGFRSMIEGFNNNPAATK-ELQAMEAOULTAMSDQLVGAD 416
DB 347 A-----ALTAENTFAIKERNENAKATYEALQYEDLAAVKAANANE 389
QY 417 GELPAEIOAKIDALQALQOPSADGLATAMGOVAFAAKVGSGSAGTAGTVQ----- 468
DB 390 ADYQAKTLAYQTELARQK-----ANMDAKAYEAAVAANNAANALTAENTAIKRN 442
QY 469 MNVKQLY--KTAESTSSSSSYAALSDGYSAYVTLNLSYSESRSGVOSATISQANPALSR 526
DB 443 ADAKADYEAKTLAYQADLAAYQKDLAD---YVKKLAYEDEOTSIIKALAE-----LEK 493
QY 527 SVRSRSGIESGGRADDSQRAEFTIVRD-----SQTIGVYSRLQYLDLSMTIYSNPQ 579
DB 494 HKNEDG-----NLTEPSAQNLYVDLEPNANLSTTGGFKASAVDADFASKSTSKAK 545
QY 580 ANOETIMQ 587
DB 546 YDQ-KILQ 552

```

RESULT 37
US-09-541-782-6
; Sequence 6, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey


```
OY 327 -----KPSGSD-----IPYGP5-----GSAASAGSAGALKSSNNSGRISLL 365
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 960 GATVSEDDGKDXDGTGTVYAKTVIDAVNKSQWVRVTEGATAEIGATAVNAGNAEYVTSCTS 1019
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 366 DDVNDEMAIAL-----QCFRSMIEQFNVNNPATAKELQAMEAQLTAMSDQLVGADGELPA 421
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1020 VFKKNGNATTATVSKDNCNINVKYDVNVGDGLKIGDDKIVADTTLT--VTGKVSYP 1077
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 422 ETQAIKDALQALKOPSDGLTAMGQVAFAPAAKVGGSAGTAGTVOMNVKQLYKTAESS 481
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1078 GANSVNN-----NKKLVNAEGLTATALNLSWTAKADKYADGESEGETDQEVKAGDKVTFFKA 1133
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 482 TSSSSYAAALSD-GYSAYKTLNLSYSESRSQVQSAISOT-----ANPALSR- 526
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1134 GKNLKVKQSEKDFYSLQDTLTGLTSTILGCTANGRNDTGYINKDGLTITLANGAAAGT 1193
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 527 -----SVSRSGIESQGRSADASQRAAETIYRDSQTLGDVYSRLQVLDLSIMSTIVSNP 578
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1194 DASNGNTISVTRKGI-SAGNKEITNVKSALKTKYKDTQNTAD-----ETQDKREFHAAYKN- 1246
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 579 QANQEEIMOKLTASIS 594
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1247 -ANEVEFVGKNGATVS 1261
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: March 12, 2003, 12:23:06
Job time : 77 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 12, 2003, 09:40:57 ; Search time 66 Seconds
(without alignments)
1332.506 Million cell updates/sec

Title: US-10-007-693-139
Perfect score: 3204
Sequence: 1 MHNNHMEGPFESVSSNQS.....PAFIQYLVNIALSFSGYLS 660

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB ID	Description
1	3204	100.0	660	22 AAU38921	C. trachomatis CT8
2	2943	91.9	647	23 ABB94283	Chlamydia trachoma
3	1690	52.7	361	21 AAB13695	Chlamydia sp. prot
4	1690	52.7	361	22 AAG83263	Protein encoded by
5	1690	52.7	361	23 ABB94234	Chlamydia protein
6	1517	47.3	331	20 AAY37572	Chlamydia trachoma
7	1435	44.8	350	20 AAY37571	Chlamydia trachoma
8	894	27.9	651	21 AAU71954	Chlamydia pneumonia
9	894	27.9	651	22 AAU38899	C. pneumoniae CT62
10	894	27.9	651	23 ABB90532	Chlamydia pneumonia

11	894	27.9	651	23 ABB94277	Chlamydia pneumonia
12	830.5	25.9	583	21 AAY71955	5'-truncated Chlam
13	453	14.1	452	21 AAY71956	3'-truncated Chlam
14	453	14.1	490	21 AAY71957	C. pneumoniae 76 k
15	451	14.1	478	20 AAY35358	Chlamydia pneumonia
16	445.5	13.9	212	20 AAY35357	Chlamydia pneumonia
17	256	8.0	6281	22 AAU37120	Staphylococcus aur
18	254	7.9	2344	22 AAU37120	Staphylococcus aur
19	251	7.8	1463	23 AAE20110	Staphylococcus aur
20	249.5	7.8	5024	22 AAG82935	Lactobacillus rham
21	248.5	7.8	5024	22 AAG82935	S. epidermidis ope
22	247	7.7	3158	22 AAU37018	Staphylococcus epi
23	244.5	7.6	2659	23 AAU75489	Staphylococcus aur
24	239.5	7.5	2434	22 AAU34339	S. aureus antigen
25	237	7.4	2137	23 ABB39618	Staphylococcus epi
26	236	7.4	1215	22 AAU34412	Staphylococcus epi
27	236	7.4	1269	22 AAU37520	Staphylococcus aur
28	235.5	7.4	2437	22 AAU34338	Staphylococcus aur
29	235	7.3	2285	20 AAM98149	Staphylococcus aur
30	234.5	7.3	1029	22 AAU34389	Bacillus subtilis
31	234.5	7.3	1048	22 AAU37490	Staphylococcus aur
32	234	7.3	560	22 AAU37464	Staphylococcus aur
33	234	7.3	1063	23 ABB54168	Lactococcus lactis
34	233	7.3	2025	22 AAU34207	Staphylococcus aur
35	228	7.1	2086	22 AAU34143	Staphylococcus aur
36	228	7.1	5795	22 AAU37017	Staphylococcus aur
37	227.5	7.1	1095	22 AAG83030	S. epidermidis ope
38	222	6.9	2478	22 AAU34320	Staphylococcus aur
39	222	6.9	2478	22 AAU37374	Staphylococcus aur
40	219.5	6.9	1833	22 ABB71141	Drosophila melanog
41	219	6.8	3696	23 ABB40235	Staphylococcus epi
42	217	6.8	2368	22 AAU34139	Staphylococcus aur
43	217	6.8	2368	22 AAU36796	Staphylococcus aur
44	214	6.7	1072	23 ABB54963	Staphylococcus aur
45	213	6.6	2186	22 AAU37320	Lactococcus lactis

ALIGNMENTS

RESULT 1	
AAU38921	
ID	AAU38921 standard; Protein: 660 AA.
XX	
AC	AAU38921;
XX	
DT	16-JAN-2002 (first entry)
XX	
DE	C. trachomatis CT875 protein.
XX	
KW	Chlamydia; sexually transmitted disease; PID; antibacterial;
KW	pelvic inflammatory disease; antigen; trachoma; gynecological;
KW	acute respiratory tract infection; atherosclerosis; male infertility
KW	coronary heart disease.
XX	
OS	Chlamydia trachomatis serovar E.
XX	
PN	MO200181379-A2.
PD	
XX	
PD	01-NOV-2001.
XX	
PF	23-APR-2001; 2001WO-US13081.
XX	
PR	21-APR-2000; 2000US-198853P.
PR	20-JUL-2000; 2000US-219752P.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Bhatia A, Probst P, Stromberg EJ;
XX	
DR	WPI; 2001-616771/71.
DR	N-PSDB; AAS56995.
XX	

PT New polynucleotide for treating Chlamydia infections encodes a
PT polynucleotides containing an immunogenic portion of a Chlamydia
PT antigen -
PS
PS Claim 3: Page 205-207; 208pp; English.
XX
XX The invention relates to isolated polynucleotide encoding at least
CC a partial Chlamydia protein which is an antigenic fragment, or the
CC complements, fragments, homologues and variants, and antibodies
CC raised against the antigenic proteins (or fragments). The nucleic
CC acids, proteins and antibodies are used to diagnose and treat Chlamydia
CC infections (e.g. a sexually transmitted disease, pelvic inflammatory
CC disease (PID), acute respiratory tract infection, trachoma,
CC atherosclerosis and coronary heart disease) in a patient, and in
CC the treatment of male infertility. The compounds of the invention are
CC also useful for detecting the presence of Chlamydia in a patient, and
CC stimulating and/or expanding T cells specific for a Chlamydia protein.
CC The present sequence represents a Chlamydia antigen.
XX
SQ Sequence 660 AA:
Query Match 100.0%; Score 3204; DB 22; Length 660;
Best Local Similarity 100.0%; Pred. No. 4.3e-185;
Matches 660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MHHHHHMEGPESSVSSNMPININGQIASNSETKESTKASEASPSASSVSMSFSL 60
DB 1 MHHHHHMEGPESSVSSNMPININGQIASNSETKESTKASEASPSASSVSMSFSL 60
OY 61 SAKNALISLRDALINKNSPTDLSOLEASTSTYTRAAKYDEAKSNFTAKSGLEN 120
DB 61 SAKNALISLRDALINKNSPTDLSOLEASTSTYTRAAKYDEAKSNFTAKSGLEN 120
OY 121 AKTAEYERKMDLMAALDMERLANSPPSNHTEEVNNTKALEOKDTIDKLNKLVTL 180
DB 121 AKTAEYERKMDLMAALDMERLANSPPSNHTEEVNNTKALEOKDTIDKLNKLVTL 180
OY 181 QONKSLTEVLKTTDSADQIPAINSOLEINKNSADQIKDLERONTSYEAVLTNAGEVIR 240
DB 181 QONKSLTEVLKTTDSADQIPAINSOLEINKNSADQIKDLERONTSYEAVLTNAGEVIR 240
OY 241 ASSEAEIKIGQALQSIYDAGDOSQAAYVLAQOONNSPDNIAATKELIDAAETKYNELKOEI 300
DB 241 ASSEAEIKIGQALQSIYDAGDOSQAAYVLAQOONNSPDNIAATKELIDAAETKYNELKOEI 300
OY 301 TGLTDSPLVKKAEQISOAKODIQEIKPSSGSDPIVGPSSAASAGALSSNNSGR 360
DB 301 TGLTDSPLVKKAEQISOAKODIQEIKPSSGSDPIVGPSSAASAGALSSNNSGR 360
OY 361 ISLLDDVDNEMAAIALQGRSMIEQFNVNPNPATAKELQAMEAQLTAMSDQLVAGDELP 420
DB 361 ISLLDDVDNEMAAIALQGRSMIEQFNVNPNPATAKELQAMEAQLTAMSDQLVAGDELP 420
OY 421 AETQAIKDALQALKOPSSADGLTAMGOYAFAAKVGGSAGTAGYVONNVKQLYTAFS 480
DB 421 AETQAIKDALQALKOPSSADGLTAMGOYAFAAKVGGSAGTAGYVONNVKQLYTAFS 480
OY 481 STSSSSYAALSDGYSAYKTLNLSYESRSGVOSAIQOTANPALSSVSSGIESQGRSA 540
DB 481 STSSSSYAALSDGYSAYKTLNLSYESRSGVOSAIQOTANPALSSVSSGIESQGRSA 540
OY 541 DASQRAAETIVRDSQTLGDIYSRLQVLDLSMTSTIVSNPQANDEIMOKLTASISKAPOFG 600
DB 541 DASQRAAETIVRDSQTLGDIYSRLQVLDLSMTSTIVSNPQANDEIMOKLTASISKAPOFG 600
OY 601 YPAVONSADSLQKFAAQOLEREFVDGERSLAESOENAFKROPAPITQOVLVNIASLFSGYLS 660
DB 601 YPAVONSADSLQKFAAQOLEREFVDGERSLAESOENAFKROPAPITQOVLVNIASLFSGYLS 660
RESULT 2
ABB94283
ID ABB94283 standard; Protein; 647 AA.

XX ABB94283;
AC 05-JUN-2002 (first entry)
XX
DT Chlamydia trachomatis protein sequence SEQ ID NO:436.
XX
DE Chlamydia trachomatis protein sequence SEQ ID NO:436.
XX
XX Chlamydia trachomatis protein sequence SEQ ID NO:436.
KW Chlamydia trachomatis protein sequence SEQ ID NO:436.
KW Chlamydia trachomatis protein sequence SEQ ID NO:436.
XX Chlamydia trachomatis protein sequence SEQ ID NO:436.
OS Chlamydia trachomatis protein sequence SEQ ID NO:436.
PN Chlamydia trachomatis protein sequence SEQ ID NO:436.
XX Chlamydia trachomatis protein sequence SEQ ID NO:436.
PD Chlamydia trachomatis protein sequence SEQ ID NO:436.
XX Chlamydia trachomatis protein sequence SEQ ID NO:436.
PF Chlamydia trachomatis protein sequence SEQ ID NO:436.
PR 20-JUL-2001; 2001WO-US23121.
PR 20-JUL-2000; 2000US-0620412.
PR 23-APR-2001; 2001US-0841132.
XX (CORI-) CORIXA CORP.
PA Filing SP, Skeiky YAM, Probst P, Bhatia A;
PI WPI; 2002-179901/23.
XX Novel compositions comprising Chlamydia Cap1 protein and its use in the
PT treatment of Chlamydia infection -
PS Claim 34: Page 391-393; 537pp; English.
XX
XX The present invention describes compositions comprising a Chlamydia Cap1
CC protein and methods for the diagnosis and therapy of Chlamydia infection.
CC Chlamydia DNA and protein sequences from the present invention can have
CC antibacterial and immunostimulant activities, and can be used in
CC vaccines. Compounds from the present invention can be used for eliciting
CC an immune response, specifically stimulating a Chlamydia-specific T-cell
CC response or inhibiting the development of a Chlamydia infection in an
CC animal. Methods from the present invention can be used for detecting the
CC presence of Chlamydia in a patient; to stimulate and/or expand T cells
CC specific for a Chlamydia protein; and for treatment of a Chlamydia
CC infection. ABB942394 to ABB942709 and ABB94096 to ABB94374 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 647 AA:
Query Match 91.9%; Score 2943; DB 23; Length 647;
Best Local Similarity 93.9%; Pred. No. 2.3e-169;
Matches 613; Conservative 15; Mismatches 19; Indels 6; Gaps 1;
OY 8 MESGPESSVSSNMPININGQIASNSETKESTKASEASPSASSVSMSFSLSSAKNALI 67
DB 1 MESGPESSVSSNMPININGQIASNSETKESTKASEASPSASSVSMSFSLSSAKNALI 67
OY 68 SLRDALINKNSPTDLSOLEASTSTYTRAAKYDEAKSNFTAKSGLENAKTAEY 127
DB 61 SLRDALINKNSPTDLSOLEASTSTYTRAAKYDEAKSNFTAKSGLENAKTAEY 120
OY 128 ETKMADLMAALQDMERLANSPPSNHTEEVNNTKALEOKDTIDKLNKLVTLQONKSL 187
DB 121 ETKMADLMAALQDMERLANSPPSNHTEEVNNTKALEOKDTIDKLNKLVTLQONKSL 174
OY 188 TEVLKTTDSADQIPAINSOLEINKNSADQIKDLERONTSYEAVLTNAGEVIRKASEAGI 247
DB 175 KETLTTTDSADQIPAINSOLEINKNSADQIKDLERONTSYEAVLTNAGEVIRKASEAGI 234
OY 248 KIGQALQSIYDAGDOSQAAYVLAQOONNSPDNIAATKELIDAAETKYNELKOEHTGTLTDS 307
DB 235 KIGQALQSIYDAGDOSQAAYVLAQOONNSPDNIAATKELIDAAETKYNELKOEHTGTLTDS 294
OY 308 LVKKAEEQISOAKODIQEIKPSSGSDPIVGPSSAASAGALSSNNSGRISLLDD 367


```

|||||
Db 295 LVKKAEQISQAQKDIOEIKPSSGSDIPYVPGSSAASASAGALKSSNNSGRISILLDD 354
OY 368 VNEMAAIALOGFRSMIEQFNVPNNPATAKEIQAMEAQLTAMSDQLVGADGELPAEIQAK 427
Db 355 VNEMAAIALMOGFRSMIEQFNVPNNPATAKEIQAMEAQLTAMSDQLVGADGELPAEIQAK 414
OY 428 DALAALQKPSADGLATAMGQVAFAPAAKVGGSAGTACTVOMNVPKOLYTAFSSSTSSSY 487
Db 415 DALAALQKPSADGLATAMGQVAFAPAAKVGGSAGTACTVOMNVPKOLYTAFSSSTSSSY 474
OY 468 AAALSDGYSAYKTLNLSYSESRGVSQSAISQTPANPALSRVSRSRGIESQGRSADASQRAA 547
Db 475 AAALSDGYSAYKTLNLSYSESRGVSQSAISQTPANPALSRVSRSRGIESQGRSADASQRAA 534
OY 548 ETVIVDSQTLGIVSRLOYLVDLSMTSTIVSNPQANOEIMOKLTASISKAPQGYPAVONS 607
Db 535 ETVIVDSQTLGIVSRLOYLVDLSMTSTIVSNPQANOEIMOKLTASISKAPQGYPAVONS 594
OY 608 ADSLQKFAQLEREFVDSERSLAESQENAFRRKQPAFIQOVLVNIASLFSGYLS 660
Db 555 ADSLQKFAQLEREFVDSERSLAESQENAFRRKQPAFIQOVLVNIASLFSGYLS 647

RESULT 3
AAB13695
ID AAB13695 standard: Protein: 361 AA.
XX
AC AAB13695:
XX
DT 02-FEB-2001 (first entry)
XX
DE Chlamydia sp. protein # 6.
XX
KW Chlamydial infection; sexually transmitted disease;
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
KW trachoma; blindness; acute respiratory tract infection;
KW atherosclerosis; coronary heart disease; antibacterial.
XX
OS Chlamydia sp..
XX
PN WO200034483-A2.
XX
PD 15-JUN-2000.
XX
PE 08-DEC-1999; 99WO-US29012.
XX
PR 08-DEC-1998; 98US-0208277.
PR 08-APR-1999; 99US-0288594.
PR 01-OCT-1999; 99US-0410568.
PR 22-OCT-1999; 99US-0426571.
XX
PA (CORI-) CORIXA CORP.
PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
XX WPI: 2000-43303/37.
XX
PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
PT comprises immunogenic portion of Chlamydia antigen, which comprises
XX amino acid sequence encoded by polynucleotide sequence -
XX
PS Disclosure: Pages 251-253; 256pp; English.
XX
CC The present invention relates to new nucleic acid sequences and the
CC proteins encoded by the nucleic acid sequences. The encoded proteins
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
CC proteins are useful for the serodiagnosis and treatment of Chlamydia
CC infection. Chlamydiae are intracellular bacterial pathogens that are
CC responsible for a wide variety of human infections. C. trachomatis
CC infection is one of the most common sexually transmitted diseases and can
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
CC and infertility. Trachoma due to ocular infection with C. trachomatis is

```

```

CC the leading cause of preventable blindness worldwide. C. pneumonia is a
CC major cause of acute respiratory tract infections in humans and is also
CC thought to play a role in the pathogenesis of atherosclerosis and
CC coronary heart disease. The present sequence is a protein isolated in the
CC present invention.
XX
SQ Sequence 361 AA:
Query Match 52.7%; Score 1690; DB 21; Length 361.
Best Local Similarity 97.0%; Pred. No. 3,9e-94;
Matches 350; Conservative 4; Mismatches 7; Indels 0; Gaps 0:
OY 300 HTGLDPSPLVKKAEQISQAQKDIOEIKPSSGSDIPYVPGSSAASAGALKSSNNSG 359
Db 1 HOEIDSPLYVKKAEQINQAOODIOTITPSGLDIPYVPGSSAASAGALKSSNNSG 60
OY 360 RISLLLDVDNEMAAIALOGFRSMIEQFNVPNNPATAKEIQAMEAQLTAMSDQLVGADGEL 419
Db 61 RISLLLDVDNEMAAIALOGFRSMIEQFNVPNNPATAKEIQAMEAQLTAMSDQLVGADGEL 120
OY 420 PAEIQAIKDALQALQKPSADGLATAMGQVAFAPAAKVGGSAGTACTVOMNVPKOLYTAF 479
Db 121 PAEIQAIKDALQALQKPSADGLATAMGQVAFAPAAKVGGSAGTACTVOMNVPKOLYTAF 180
OY 480 SSTSSSYAAALSDGYSAYKTLNLSYSESRGVSQSAISQTPANPALSRVSRSRGIESQGRS 539
Db 181 SSTSSSYAAALSDGYSAYKTLNLSYSESRGVSQSAISQTPANPALSRVSRSRGIESQGRS 240
OY 540 ADASQRAETIVRDSQTLGDVYSRLQYLDLSMTSTIVSNPQANOEIMOKLTASISKAPQF 599
Db 241 ADASQRAETIVRDSQTLGDVYSRLQYLDLSMTSTIVSNPQANOEIMOKLTASISKAPQF 300
OY 600 GYPVONSNDSIQKFAQLEREFVDSERSLAESQENAFRRKQPAFIQOVLVNIASLFSGYL 659
Db 301 GYPVONSNDSIQKFAQLEREFVDSERSLAESQENAFRRKQPAFIQOVLVNIASLFSGYL 360
OY 660 S 660
Db 361 S 361

RESULT 4
AAG83263
ID AAG83263 standard: Protein: 361 AA.
XX
AC AAG83263:
XX
DT 05-SEP-2001 (first entry)
XX
DE Protein encoded by Chlamydia trachomatis clone CT622.
XX
KW Chlamydia; vaccine; infection; fusion protein; antigen;
KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
KW acute respiratory tract infection; Cap1; CF529; OMCB;
KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.
XX
OS Chlamydia trachomatis.
XX
PN WO200140474-A2.
XX
PD 07-JUN-2001.
XX
PE 04-DEC-2000; 2000WO-US32919.
XX
PR 03-DEC-1999; 99US-0454684.
PR 19-APR-2000; 2000US-0556877.
PR 20-JUN-2000; 2000US-0598419.
XX
PA (CORI-) CORIXA CORP.
PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;
XX WPI: 2001-374831/39.

```


Dd		307	LVKKAEEQINQAQQDIDQTTPSPGLDIPIVGPSCS	340
	RESULT 8			
ID	AAY71954			
XX	AAY71954 standard; Protein; 651 AA.			
AC	AAY71954;			
DT	26-MAR-2001 (first entry)			
XX				
DE	Chlamydia pneumoniae 76 kDa full-length protein.			
XX				
KW	76 kDa protein; bactericidal; diagnosis; prevention; treatment; pneumonia; upper respiratory tract disease; bronchitis; sinusitis; acute respiratory disease; cough; sore throat; hoarseness; fever; vaccine; immunisation.			
XX				
OS	Chlamydia pneumoniae.			
XX				
PX	WO200066739-A2.			
PD	09-NOV-2000.			
PE	03-MAY-2000; 2000WO-CA00511.			
PR	03-MAY-1999; 99US-0132270. 30-JUN-1999; 99US-0141276.			
PA	(AVET) AVENTIS PASTEUR LTD.			
PI	Muridin AD, Oomen RP, Wang J, Dunn P; WPI: 2000-687542/67. DR N-PADB: AAD02063.			
PT	Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae, useful for vaccinating against Chlamydia infections - Claim 16a; Fig I; 112pp; English.			
PS				
CC	The present sequence is Chlamydia pneumoniae full-length 76 kDa protein. C. pneumoniae 76 kDa protein is used in the diagnosis, prevention and treatment of C. pneumoniae infections (e.g. pneumonia, upper respiratory tract disease, bronchitis, sinusitis and acute respiratory disease such as cough, sore throat, hoarseness, fever; and abnormal chest sounds on auscultation). C. pneumoniae sequence is also used as vaccines for immunising humans against diseases caused by C. pneumoniae.			
CC				
SQ	Sequence 651 AA:			
Query Match	27.9%; Score 894; DB 21; Length 651;			
Best Local Similarity	34.6%; Pred. No. 8.4e-46;			
Matches 233;	Conservative 126; Mismatches 262; Indels 52; Gaps 17,			
OY	11 GPESVSNSNDSNNPILINGIASNSEKTEKA-----SEASP--SASSSVSGWSFLSKAK 63			
I :	: : : : : : : : : : : : : : : : : : : : : :			
Dd	6 GGPIDETEREPAPLDAAGLEASANKAEQRIRIAGAFAEKPKESKTDSVERMSIIIRSAV 65			
OY	64 NLALSLRD--ATLNKNSSPTDSLQSULEASTSTSYTRVAIAKDYEAKSNFDTAKGLENA 121			
:	: :			
Dd	66 NAIMSLADKLGIASSNSSSTRSR--ADVDDTTAATAAPPFPFDOKTTOAQTAHYDTIFTS 124			
OY	122 KTLAAEFTKMADLMALAODMERLANSDPSNNHTEEVANNIKALEAQDKTDID-----K 173			
I :	: : : : : : : : : : : : : : : : : : : : : : : :			
Dd	125 TSLADIQAALVALSELQAVNTMIKPFTAATD-----EF----ALAAEWERTKNDLVKGGAQ 173			
OY	174 LMKLYTVLONQNLSLEVLTCTDSAOIPAIINSQLEFNKKNSAQOIKDLERONI---SYEA 230			
I :	: : : : : : : : : : : : : : : : : : : : : : : :			
Dd	174 ITELKVASDNQALIDSLECKLTFPLQLGALLQSVANNNKAELLKEQMDNPVVDPKTPA 233			

OY	231	VLIHNGEYIKRSSEGGITLUGALOSIYDAGQSOAAVLQAOQNNSPNIIATKELIDAE	230
OY	231	VLIHNGEYIKRSSEGGITLUGALOSIYDAGQSOAAVLQAOQNNSPNIIATKELIDAE	230
Db	234	IAQSLVDPDTATTAQIQIENDDGNAIRDAYFAGQNASGAVENAKSNNSISNIDSAKALITAK	293
OY	291	TKVNLKOEHNGLMTSPILPKKAEQISOAKDIOEIKPS-GSDPIYGPS-GSAASAGSA	348
Db	294	TOIAFAOKK---FPDSPILOAEQWVIOAEKDLNKRIPADGSDVPPNCGTYVGGSKQOGSS	350
OY	349	AGALKSNNNSGRISILLDDVDNEMAAIALQGRSMIDQFVYNNPATAKELQAMEAQLTAM	408
Db	351	IGSI-----RVSMILLDAENETASILMSGFRQIMHFTENTENDPSQAQOELAAQARAA	403
OY	409	SDQLVAGDGLPAELQAIKDLAOL-KQPSADGLATAMQVAPAAAKVGGSGAGTAGTV	467
Db	404	K---AAGDSSAAALADQKQALKALGAKGQOQIILNALGQIASAAVYSAVGPPAAASSI	460
OY	468	QMNWOLKKTAFSSSTSSSYAALSDGSAVKTNTLSYSESR-GVQSAISQTNPALSR	526
Db	461	GSSVQVLKKT--SKSTGSDYTTQISAGTDAYKSIINDAYGARANDATDVINNSTPLTR	518
OY	527	SVSRGSIQSGRSADASORAETIVRDSOTLGDVYSRLQVLDLSIMSTIVSNPQANDEIM	586
Db	519	SVPRARTEARC-PEKTDQALARVISGNSRTIGDYVSQVSAIQSVMTQIQSNPQANNEIR	577
OY	567	OKLTASISKAPQFCYPAYONSAISLQKFAQLEREPEYDGERSLAESQENAFKQPAFIQO	644
Db	578	OKLTSAVTKPPQFGPYVOLSDNSTOKFIATLESLEFAEGSRTAETIKALSFTNSLFTQO	637
OY	647	VLVNTASTLFGYTL 659	
Db	638	VLVNTGSLYSGL 650	
RESULT 9			
AAU38899			
ID	AAU38899	standard; Protein; 651 AA.	
XX	AAU38899;		
AC	16-JAN-2002	(first entry)	
XX			
DT			
DE	C. pneumoniae CT622	homologue Cpn0728.	
XX			
KW	Chlamydia; sexually transmitted disease; PID; antibacterial;		
KW	pelvic inflammatory disease; antigen; trachoma; gynecological;		
KW	acute respiratory tract infection; atherosclerosis; male infertility;		
KW	coronary heart disease.		
OS	Chlamydia pneumoniae.		
XX			
PN	WO200181379-A2.		
XX			
PD	01-NOV-2001.		
XX			
PF	23-APR-2001; 2001WO-US13081.		
XX			
PR	21-APR-2000; 2000US-198853P.		
PR	20-JUL-2000; 2000US-219752P.		
XX			
PA	(CORI-) CORIXA CORP.		
XX			
PI	Bhatia A, Probst P, Stromberg EJ;		
DR	WPI: 2001-616771/71.		
XX	N-PSDB; AAS57031.		
XX			
PT	New polynucleotide for treating Chlamydia infections encodes a		
PT	polynucleotides containing an immunogenic portion of a Chlamydia		
XX	antigen -		
PS	Disclosure: Page 161-162; 208pp; English.		
CC	The invention relates to isolated polynucleotide encoding at least		

CC a partial Chlamydia protein which is an antigenic fragment, or the
 CC complements, fragments, homologues and variants, and antibodies
 CC raised against the antigenic proteins (or fragments). The nucleic
 CC acids, proteins and antibodies are used to diagnose and treat Chlamydia
 CC infections (e.g. a sexually transmitted disease, pelvic inflammatory
 CC disease (PID), acute respiratory tract infection, trachoma,
 CC atherosclerosis and coronary heart disease) in a patient, and in
 CC the treatment of male infertility. The compounds of the invention are
 CC also useful for detecting the presence of Chlamydia in a patient, and
 CC stimulating and/or expanding T cells specific for a Chlamydia protein.
 CC The present sequence represents a Chlamydia antigen.

XX Sequence 651 AA;

Query Match 27.9%; Score 894; DB 22; Length 651;
 Best Local Similarity 34.6%; Pred. No. 8.4e-46;

Matches 233; Conservative 126; Mismatches 262; Indels 52; Gaps 17;

QY 11 GPESVSSNOSSMNPINQIANSSETKSTKA-----SEASP--SASSSVSWSPFLSSAK 63
 || :
 Db 6 GGPIDETERPPADLSAQGLEASANKSAEORAGAEAPKESKTSVERWSTLRSAV 65
 QY 64 NALISLRD-AI LKNNSSPTDSLQLEASTSTSTVTRVAANDYDEAKSNPTAKSGLENA 121
 || :
 Db 66 NALMSLADKLGIASSNSSSTSRs-ADVSTTATAPTPPPFDYKTKQAQATAYDTIFTS 124
 QY 122 KTLAEYETKMDLMAALODMERLANSNHTTEEVNNIKALEAKOKTID-----K 173
 || :
 Db 125 TSLADIQALVSLQDAVINIKDTAATD-----EET-----AIAEWETKNAADAVKGAQ 173
 QY 174 LNKLTTLQONKSLTEVLTJKTSDADQIPAINSOLEINKNSADQIKLEERQNI---SYEA 230
 || :
 Db 174 ITELAKYASDQALITDSLKLTSPDLQALLOAVANNKKAELKEQNDPVYVGKTPA 233
 QY 231 VLTNAGEVIKASSEAGIKLQALOSIYDAGDQSAVLAQOQNSPNMIATKELIDAE 290
 || :
 Db 234 IAOQSLVDTDAATQIEKQGNAIRDAYFAGQNASGAVENAKSNISNIDSKAKAIATAK 293
 QY 291 TKVNELKEHNGLDSPLYVKAKEEIOSONODIOEIKPS-GSDIPIVPS-GSAASASA 348
 || :
 Db 294 TQIEAQK---FPDSPLQEHQWVIOAEKDLKNIKPADSSDVNPPTVGGSKQGGSS 350
 QY 349 AGALKSSNNSGRIISLLDDVDNEMAAILOGFSMIEQFNVNPNPTAKELQAMEKQLTRAM 408
 || :
 Db 351 IGSIT-----RVSMILDDAENETASTILMSGFRQIMHFNFENPDSQAQOELAQARAA 403
 QY 409 SDQLVGADGELPAETIOAIKDALAQL-KQPSADGIATMGQVAFAPAAKVGGSAGTACTV 467
 || :
 Db 404 K---AAGDSSAAALADQAKALEALGKAGQOOGITLNLGQIASAAVYAGVPAASAASI 460
 QY 468 QMNVKQLTKTAFSSSTSSSYAAALSDGYSAKYKTLNLSYSEKS-GVQSAISGTANPALSR 526
 || :
 Db 461 GSSVQVLKTT--SKSTGSDYKTQISAGYDAKSIDAYGARANDTRVINNVSPALTR 518
 QY 527 SVRSRGISQGRSADASORAAETIYRDSQTLGDVYSRLQVDSLSTIVSNQOANQEIM 586
 || :
 Db 519 SVPARTEARG-PEKTDQALARVISGNSRTIGDYVSQVSALOSVOIIOISNPQANNEIR 577
 QY 587 OKLTASISKAPQFGYPAVONSADSLQKFAQOLEREYVDGERSLAESQENAFRKQAPATQ 646
 || :
 Db 578 OKLTASAVYKPPQFGYPAVQLSNDSTQKFIKLESIFAEGSRTAAEIKALSFETNSLFIQ 637
 QY 647 VLVNTASLFGSYL 659
 || :
 Db 638 VLVNIGSLYSGYL 650

RESULT 10

ABB90532

ID ABB90532 standard; Protein: 651 AA.

XX ABB90532;

XX

DT 29-JUL-2002 (first entry)

XX Chlamydia pneumoniae cp7033 protein, SEQ ID NO:13.

XX Chlamydial infection; antigen; immunogen; vaccine; diagnosis;

KW human respiratory disease; cardiovascular disease; atherosclerosis;

KW coronary artery disease; carotid artery stenosis; myocardial infarction;

KW cerebrovascular disease; aortic aneurysm; claudication; stroke;

XX strain CML029.

XX Chlamydia pneumoniae.

XX WO200202606-A2.

XX 10-JAN-2002.

XX 03-JUL-2001; 2001WO-IB01445.

XX 03-JUL-2000; 2000GB-0016363.

XX 11-JUL-2000; 2000GB-0017047.

XX 21-JUL-2000; 2000GB-0017983.

XX 07-AUG-2000; 2000GB-0019368.

XX 18-AUG-2000; 2000GB-0020440.

XX 14-SEP-2000; 2000GB-0022583.

XX 10-NOV-2000; 2000GB-0027549.

XX 22-DEC-2000; 2000GB-0031706.

XX (CHIR-) CHIRON SPA.

XX Ratti G, Grandi G;

XX WPI: 2002-154726/20.

XX N-PSDB: ABL91190.

XX Novel Chlamydia pneumoniae protein useful in the manufacture of a

XX medicament for treatment or prevention of infection due to Chlamydia,

XX preferably Chlamydia pneumoniae, and for diagnostic purposes -

XX Claim 1: Page 47; 364pp: English.

XX Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia

XX pneumoniae (strain CML029), and ABL91184-ABL91373 represent DNA encoding

XX them. The proteins are predicted to be immunogenic and may therefore be

XX useful in vaccine production and for diagnostic purposes. Chlamydia

XX pneumoniae is a common cause of respiratory disease in humans, and is

XX also involved in the development of cardiovascular diseases such as

XX atherosclerosis, coronary artery disease, carotid artery stenosis,

XX myocardial infarction, cerebrovascular disease, aortic aneurysm,

XX claudication and stroke. The proteins and nucleic acids of the invention

XX may be used in vaccines and pharmaceutical compositions for the

XX prevention or treatment of chlamydial infections, particularly Chlamydia

XX pneumoniae infections. The proteins may also be used in the detection of

XX Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched

XX DNA probe assay or blotting techniques for determining Chlamydia

XX pneumoniae gene expression. The present sequence represents a

XX specifically claimed Chlamydia pneumoniae protein of the invention.

XX Sequence 651 AA;

Query Match 27.9%; Score 894; DB 23; Length 651;

Best Local Similarity 34.6%; Pred. No. 8.4e-46;

Matches 233; Conservative 126; Mismatches 262; Indels 52; Gaps 17;

QY 11 GPESVSSNOSSMNPINQIANSSETKSTKA-----SEASP--SASSSVSWSPFLSSAK 63

Db 6 GGPIDETERPPADLSAQGLEASANKSAEORAGAEAPKESKTSVERWSTLRSAV 65

QY 64 NALISLRD-AI LKNNSSPTDSLQLEASTSTSTVTRVAANDYDEAKSNPTAKSGLENA 121

Db 66 NALMSLADKLGIASSNSSSTSRs-ADVSTTATAPTPPPFDYKTKQAQATAYDTIFTS 124

QY 122 KTLAEYETKMDLMAALODMERLANSNHTTEEVNNIKALEAKOKTID-----K 173

|| :

```
DB 125 TSLADIQAALVLSQDAVYNIKDTAATD-----EET-----AIAEMWETKNADAVKGAQ 173
OY 174 LNKLVTLQONKSLTEVLTQTTSDAQIIPAINSQLEINKNSADQIINKDELRONI---SYEA 230
DB 174 ITELAKYASDNQAIIDSLSKLTISFDLQALLOSVAANNKKAELLEMQDNVYPPKTRA 233
OY 231 VLTNAGEVIKASSEAGIKLQALQSITVDAGDOSQAALVLAQOONSPDNIAATKELIDAAE 290
DB 234 IQOSLVQDQDATATQTEKGNMARIADYFAGQNASGAVENAKSNSISNIDSAKAALATATK 293
OY 291 TKYNELKQHTGLTDSPLVKKAEQISOAKQDIQELKPS-GSDIPIVGPS-GSASAGSA 348
DB 294 IQOIAEQAOKK---PPDSPILQEAQWYIOAEKDKLNKIPADGSDVPNPGTIVGSKQOGSS 350
OY 349 AGALKSSNNSGRISLLDDVDNEMAAIALQGFMSITEOFNVNPNPAKLEQAEQULTAM 408
DB 351 IGS1-----RYSMLDDAENETASILMSGFRQMIHMTENTEPDQAAQOELAAQARA 403
OY 409 SDQLVGADGELPAEIQAIKDALAQAL-KOPSADGLATAMGOVAFAAKVGGSAGTAGTV 467
DB 404 K---AAGDSSAAALADQAQKALEALGKAGQOQGIINALQOIASAAVVSAGVPPAAASST 460
OY 468 QMNVKQLYTAPSTSSSSSYAAALSDGYSAKYKTLNLSYSESR-GVQSAISQTNALSR 526
DB 461 GSSVKQLYKT--SKTGSQYKTQISAGYDAYKXSINDAYGARADARVDYINNVSTPALTR 518
OY 527 SVSRSGIESQGRSADASQRAAEITVDSQTLGDVYSRLQVLDSLMSTIYSPQANQEEIM 586
DB 519 SVPRATEARNG-PEKTDQALARYISGNSRKLGDVYSQVALSQQVMQIISNPQANNEEIR 577
OY 587 OKLTASISKAPQFGYPAVONSADSLQKFAAOLEREFEVDERSIAESQENAFKROPATQO 646
DB 578 OKLTSAVTKPPQGGYVYVOLSNDSTQKFTAKLESLEFAGESRTAAELKALSFETNSLFIQO 637
OY 647 VLVNIASLFSGYL 659
DB 638 VLVNIGSLYSGYL 650

RESULT 11
ABB94277
ID ABB94277 standard; Protein; 651 AA.
XX
AC ABB94277;
XX
DT 05-JUN-2002 (first entry)
XX
DE Chlamydia pneumoniae protein sequence SEQ ID NO:405.
XX
KW Chlamydial infection; Chlamydia; vaccine; detection; diagnosis;
KW antigen; antibacterial; immunostimulant; immune response;
KW Chlamydia-specific T-cell response.
XX
OS Chlamydia pneumoniae.
XX
PN WO200208267-A2.
XX
PD 31-JAN-2002.
XX
PF 20-JUL-2001; 2001WO-US23121.
XX
PR 20-JUL-2000; 2000US-0620412.
XX
PR 23-APR-2001; 2001US-0841132.
XX
PA (CORI-) CORIXA CORP.
XX
PI Fling SP, Skeiky YAM, Probst P, Bhatia A;
XX
DR WPI; 2002-179901/23.
XX
PT Novel compositions comprising Chlamydia CapI protein and its use in the
XX treatment of Chlamydia infection -
```

```
PS Disclosure; Page 370-372; 537pp; English.
XX
CC The present invention describes compositions comprising a Chlamydia CapI
CC protein and methods for the diagnosis and therapy of Chlamydia infection.
CC Chlamydia DNA and protein sequences from the present invention can have
CC antibacterial and immunostimulant activities, and can be used in
CC vaccines. Compounds from the present invention can be used for eliciting
CC an immune response, specifically stimulating a Chlamydia-specific T-cell
CC response or inhibiting the development of a Chlamydia infection in an
CC animal. Methods from the present invention can be used: for detecting the
CC presence of Chlamydia in a patient; to stimulate and/or expand T cells
CC specific for a Chlamydia protein; and for treatment of a Chlamydia
CC infection. AB192394 to AB192709 and ABB94096 to ABB94374 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 651 AA.

Query Match 27.9%; Score 894; DB 23; Length 651;
Best Local Similarity 34.6%; Pred. No. 8,4e-46;
Matches 233; Conservative 126; Mismatches 262; Indels 52; Gaps 17;

OY 11 GPESVSSNOSNMPIINGQIASNSETKETKA-----SPASP--SASSSVSSMSFLSSAK 63
DB 6 GPEPIDETERPPADLSAQGLESAANKSAEQRAGAEAKRESKTSQVSERSILRSVAV 65
OY 64 NLILSLRD--ALINKNSPTDLSQLEASTSTVYTRVAAKYDEAKSNFDTAKSGLENA 121
DB 66 NMLMSLADLGLASSNSSSTSRSS-ADVSTATATAPPPPPFDDQYKQOATAYDTIFTS 124
OY 122 KTLAEYETMADIMAALOMERLANSQPSNHTTEVYNNIKKALEAKQDTID-----K 173
DB 125 TSLADIQAALVLSQDAVYNIKDTAATD-----EET-----AIAEMWETKNADAVKGAQ 173
OY 174 LNKLVTLQONKSLTEVLTQTTSDAQIIPAINSQLEINKNSADQIINKDELRONI---SYEA 230
DB 174 ITELAKYASDNQAIIDSLSKLTISFDLQALLOSVAANNKKAELLEMQDNVYPPKTRA 233
OY 231 VLTNAGEVIKASSEAGIKLQALQSITVDAGDOSQAALVLAQOONSPDNIAATKELIDAAE 290
DB 234 IQOSLVQDQDATATQTEKGNMARIADYFAGQNASGAVENAKSNSISNIDSAKAALATATK 293
OY 291 TKYNELKQHTGLTDSPLVKKAEQISOAKQDIQELKPS-GSDIPIVGPS-GSASAGSA 348
DB 294 IQOIAEQAOKK---PPDSPILQEAQWYIOAEKDKLNKIPADGSDVPNPGTIVGSKQOGSS 350
OY 349 AGALKSSNNSGRISLLDDVDNEMAAIALQGFMSITEOFNVNPNPAKLEQAEQULTAM 408
DB 351 IGS1-----RYSMLDDAENETASILMSGFRQMIHMTENTEPDQAAQOELAAQARA 403
OY 409 SDQLVGADGELPAEIQAIKDALAQAL-KOPSADGLATAMGOVAFAAKVGGSAGTAGTV 467
DB 404 K---AAGDSSAAALADQAQKALEALGKAGQOQGIINALQOIASAAVVSAGVPPAAASST 460
OY 468 QMNVKQLYTAPSTSSSSSYAAALSDGYSAKYKTLNLSYSESR-GVQSAISQTNALSR 526
DB 461 GSSVKQLYKT--SKTGSQYKTQISAGYDAYKXSINDAYGARADARVDYINNVSTPALTR 518
OY 527 SVSRSGIESQGRSADASQRAAEITVDSQTLGDVYSRLQVLDSLMSTIYSPQANQEEIM 586
DB 519 SVPRATEARNG-PEKTDQALARYISGNSRKLGDVYSQVALSQQVMQIISNPQANNEEIR 577
OY 587 OKLTASISKAPQFGYPAVONSADSLQKFAAOLEREFEVDERSIAESQENAFKROPATQO 646
DB 578 OKLTSAVTKPPQGGYVYVOLSNDSTQKFTAKLESLEFAGESRTAAELKALSFETNSLFIQO 637
OY 647 VLVNIASLFSGYL 659
DB 638 VLVNIGSLYSGYL 650

RESULT 12
AA171955
ID AA171955 standard; Protein; 583 AA.
```

```

XX AAY71955;
AC 26-MAR-2001 (first entry)
XX
XX 5'-truncated Chlamydia pneumoniae 76 kDa protein.
XX
XX 76 kDa protein: bactericidal; diagnosis: prevention;
XX pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
XX acute respiratory disease; cough; sore throat; hoarseness; fever;
XX vaccine; immunisation; treatment; truncation mutant; mutin.
XX
XX Chlamydia pneumoniae.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Misc-difference 497 /note= "Encoded by ACT"
XX
XX WO20066739-A2.
XX
XX 09-NOV-2000.
XX
XX 03-MAY-2000; 2000WO-CA00511.
XX
XX 03-MAY-1999; 99US-0132270.
XX 30-JUN-1999; 99US-0141276.
XX
XX (AVET ) AVENTIS PASTEUR LTD.
XX
XX Murdin AD, Oomen RP, Wang J, Dunn P;
XX
XX WPI: 2000-687542/67.
XX N-PSDB: AAD02064.
XX
XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX useful for vaccinating against Chlamydia infections -
XX
XX Claim 16b: Page 100-102; 112pp: English.
XX
XX
XX The present sequence is 5'-truncated
XX Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
XX protein is used in the
XX diagnosis, prevention and treatment of C. pneumoniae
XX infections (e.g. pneumonia, upper respiratory
XX tract disease, bronchitis, sinusitis and acute respiratory
XX disease such as cough, sore throat, hoarseness, fever; and
XX abnormal chest sounds on auscultation). C. pneumoniae sequence
XX is also used as vaccines for immunising humans against diseases
XX caused by C. pneumoniae.
XX
XX
XX Sequence 583 AA:
SQ
Query Match 25.9%; Score 830.5; DB 21; Length 583;
Best Local Similarity 34.9%; Pred. No. 4.9e-42;
Matches 213; Conservative 117; Mismatches 235; Indels 45; Gaps 15;
OY 67 ISLRD--AIIKNKSPFSDLSQLEASTSTSTVTRAAKDYDEAKSNFTDAKSGLENAKTL 124
DB 1 MSLADKLGIASSNSSSSSTSR--ADVSTTATAPTPTPPFEDDYKTKQAQVATPIFTSTSL 59
OY 125 AEYETKMDLMAADMDMELANSDPSSNHTTEVNNIKKALEOKTID-----KLAK 176
DB 60 ADIQALVSLQADVNIKPTAATD-----ETP-----AIAEWEKKNDAVKGAGQITE 108
OY 177 LVTLQNNKSLTEVLKTTSDADQIPAINSOLEINKNSADQIIKDERQNI---SYFAVLT 233
DB 109 LAKYASDNQAIIDSLCKLTPDLLQALQSVANNKKALELEKMDNPNVVGKTPAIHQ 168
OY 234 NGEVYKASSEAGIKLQALQSTVDAGDOSQAAYLQAGQNNSPDNIAATKELIDAETKV 293
DB 169 SLVDQTDATATQIEKDGNAIRDAVFAQGNASGAVENAKSNNSISNIDSAKAATATKTOI 228

```

```

OY 294 NELKQEHNTGLTSDPLVKKAEQISOAKDIOETKPS-GSDIPVGPS-CSAASAGSAGA 351
DB 229 AFAQKK---FPSPPIQEBQWVIOAEKDLKIKRADGSDVNPPTVGGSKQGGSSIGS 285
OY 352 LKSSNNGRIILLDDVDNEMAAIALQGRSMIEQFNWNNPATAKELQAMEAQLTAMSDQ 411
DB 286 I-----RVSMLDDAEKETASILMSGFRQMIHMFNTENPDSQAQOELAQARAAK-- 336
OY 412 LVGAGCELPALQIAIKDALQAL-KQPSADGIATAMGOVAFPAAKVGGSGAGTACTVQMN 470
DB 337 -AAGDSAAAALADLQKALEALGRAGQOQGLMGLQIASAAVVSAGVPAASSIGSS 395
OY 471 VKQLYKTAFTSPSSSYAALSDGYSAKYKTLNLSYSESPS-CVQSAISOTANPALSRVS 529
DB 396 VKQLYKT--SKSTGSDYKQISAGYDAYKSIDATGRANDTRDVYNNVSTPALTRSP 453
OY 530 RSGIESQGRSADASQRAETIVRDSQTLGDVYSRLQVLDLSMTSTVSNPQANQEEIMOKL 589
DB 454 RARTEARG-PERTDQALARVIGNSRTLDVYSQVSAQSVQIIQSNPQANNEEIROKL 512
OY 590 TASISKAPQFGYPAYQNSADSIQKFAQLEREFVQDERSLASQENAPRKQPAFTQVLY 649
DB 513 TSAVTKPPQGYVYQLSNDSTOKFTAKLESIFAQGSRTAECIKALSFTNSLFTQVLY 572
OY 650 NIASLFSGYL 659
DB 573 NIGSLXGYL 582

```

```

RESULT 13
AAY71956
ID AAY71956 standard: Protein; 452 AA.
XX
XX AAY71956;
XX
XX 26-MAR-2001 (first entry)
XX
XX 3'-truncated Chlamydia pneumoniae 76 kDa protein.
XX
XX 76 kDa protein: bactericidal; diagnosis: prevention;
XX pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
XX acute respiratory disease; cough; sore throat; hoarseness; fever;
XX vaccine; immunisation; treatment; truncation mutant; mutin.
XX
XX Chlamydia pneumoniae.
XX
XX Synthetic.
XX
XX WO20066739-A2.
XX
XX 09-NOV-2000.
XX
XX 03-MAY-2000; 2000WO-CA00511.
XX
XX 03-MAY-1999; 99US-0132270.
XX 30-JUN-1999; 99US-0141276.
XX
XX (AVET ) AVENTIS PASTEUR LTD.
XX
XX Murdin AD, Oomen RP, Wang J, Dunn P;
XX
XX WPI: 2000-687542/67.
XX N-PSDB: AAD02065.
XX
XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX useful for vaccinating against Chlamydia infections -
XX
XX Claim 16c: Page 104-106; 112pp: English.
XX
XX The present sequence is 3'-truncated
XX Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
XX protein is used in the
XX diagnosis, prevention and treatment of C. pneumoniae
XX infections (e.g. pneumonia, upper respiratory

```


OT 13-SEP-1999 (first entry)

XX

DE Chlamydia pneumoniae surface exposed polypeptide.

XX

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

XV sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KX vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN MO9927105-AZ.

PD

XX 03-JUN-1999.

XX

PF 20-NOV-1998; 98WO-IBO1830.

XX

PR 04-NOV-1998; 98US-0107078.

FR 21-NOV-1997; 97FR-0014673.

PA (GENSET) GENSET.

xx

PI Griffais R;

xx

DR WPI: 1999-357842/30.

xx

PS Genome sequence of Chlamydia pneumoniae

Page 1156: Disclosure; 1912pp; English.

AAY34584-Y35879 represent the proteins encoded by all the open reading frames in the complete genome (see AAX91990) of Chlamydia pneumoniae. C. pneumoniae causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.

```
Sequence      478 AA:
SQ
Query Match          14.1% Score 451 DB 20 Length 478;
Best Local Similarity    29.0%; Pred. No. 2.9e-19;
Matches   135; Conservative    94; Mismatches 189; Indels    48; Gaps        14;
QY       11 GPESVSNQSSNMPTINGOIANSSETEKSTKA-----SEASP--SASSSVSSWSFLSSAK     63
DB              |||::||::| | | : | : || ::||| |||::| | || ||| 
QY       8 GPGPIDETERRPPADLSAQGLEASANKSAEQAIRIAGAEAFKEKSQTDSVERMSILRSAY     67
QY       64 NALISLRP-AIALNNSSPTDSLQSLEASTISTYTRVAARDYDIAKSNEFTAKSGLERNA     121
                |||::|||::| | : | | : | : || ::||| |||::| | || ||| 
QB       68 NAALSMLACKLGIIASSNSSSTSRS-ADVDTATAPTPPPPFPFDKYKTGAQTAVDTIETS     126
QY       122 KTLAIEFKMADMALDLDMERLADSDDSNHTEEVNNIKRLAEOKDPTID-----K     173
DB               |||::| | : | : || ::||| |||::| | || ||| 
QY       127 TSLADIQAALVSLDAVINNIKATAFID-----EEI-----AIAEWETKNDAVKGHAQ     175
DB               |||::| | : | : || ::||| |||::| | || ||| 
QY       174 LNKLTLLIONOKSKLTEVLTKTSDADQIPAINSQLINKNSADOITKLDERONI---SYEA     230
DB               |||::| | : | : || ::||| |||::| | || ||| 
QY       176 ITELAKKVASDAQAILIDSGLKITSPQLLOALLQSVAANNKANMAELLEKENGDNVVGKTPA     235
QY       231 VLTNAGEVIKASSEGIGRKGALQALOSTIVADGSOCAAVILOAOQNNSPDNIATATELIDAAE     290
DB       236 IAQSLEVDPDTDATIIDTDIEDGNAIRPAYFAOGAASCAVENAKSNSNISINDSKAIATAAK     295
QY       291 TKVNELKOEHNGILTSPPLVKRAKEGIOISOAKRDIOEKFPS-GSDIFLVGPS-CSAASAGA     348
DB       296 TOIAEAQQK----EPDSPFIQEAEQNAVIOAEKDKLNINIKPADGSVDVNPGETTYGGSGQGSS     352
QY       349 AGALKSSNGRISSLLDLDVVDEMAMAIALQGRSMITEOFNVNPNPATARELAMAEALVTAM     408
                |||::| | : | : || ::||| |||::| | || ||| 
```

Dd	353	IGST-----RVSKLLDDAENETASIIISGCFROMIHFNFENTENDPSQAAQOEELAAQAARAA	405
Oy	409	SDQLVGADDELPAEIOAIKDALAQAL-KOPSADGLATAMGOVAFAPAA	453
		: : : :	
Dd	406	K---AAGDASAALADADAQKALEALGAAGQQGGITLNLGQIASANA	448
 RESULT 16 AA35357 ID AAY35357 standard; Protein: 212 AA.			
AC	AAY35357;		
XX			
DT	13-SEP-1999	(first entry)	
XX			
DE	Chlamydia pneumoniae transmembrane protein sequence.		
XX			
KW	Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;		
RW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;		
KW	vaccine; neutralising epitope.		
XX			
OS	Chlamydia pneumoniae.		
XX			
PN	MO9927105-A2.		
XX			
PD	03-JUN-1999.		
XX			
PE	20-NOV-1998;	98WO-IB01890.	
XX			
PR	04-NOV-1998;	98US-0107078.	
PR	21-NOV-1997;	97FR-0014673.	
XX			
PA	(GEST) GENSET.		
XX			
P1	Grlffals R:		
XX			
DR	WPI; 1999-357842/30.		
XX			
XX	Genome sequence of Chlamydia pneumoniae		
PT			
XX	Page 1155; Disclosure; 1912pp; English.		
PS			
CC	AAY34584-Y35879 represent the proteins encoded by all the open reading		
CC	frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.		
CC	C. pneumoniae causes respiratory disease such as pneumonia and		
CC	bronchitis and is thought to be a contributing factor in heart		
CC	disease, sarcoidosis, sinusitis, purulent otitis media, erythema		
CC	nodosum or pharyngitis. The polypeptides encoded by the open reading		
CC	frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in		
CC	immunogenic compositions as vaccines. Vectors containing C. pneumoniae		
CC	nucleotide sequences can also be used as immunogenic compositions,		
CC	especially where the vector directs the expression of a neutralising		
CC	epitope of C. pneumoniae.		
SQ			
Sequence	212 AA;		
 Query Match 13.9%; Score 445.5; DB: 20; Length 212; Best Local Similarity 47.6%; Pred. No. 2.1e-19; Matches 100; Conservative 34; Mismatches 69; Indels 7; Gaps 4;			
Oy	451	FAAAVGVGGSGTAGTYOMNVNLYKTAFSSSTSSSYAALSDGYSAVKTLNTLSYESRS	510
		: : : : : : : : : : : : : : :	
Dd	8	FYSAGVPPMAASSIS---SVNQLYKT--SKSTGSDYTTQISAGVDAYKSTINDAYGRARN	62
		: : : : : : : : : : : : : : : : :	
Oy	511	-GVQSALISOTANPALSRSSVSNSRGIESQGRSADASGRAAEITVRDSOTLGDIYSRLQVIDS	569
		: : : : : : : : : : : : : : : :	
Dd	63	DATRIVINNVTSPALTTRSVPRAREARG-PEKTDQALARVISGNRTLTGDVYSQVSALOS	121
		: : : : : : : : : : : : : : : : :	
Oy	570	LMSITVSNPQANOGEIMOKLTASTISKAPQFCYPANQNADSLOKRAAQLEREPUJGERSL	629
		: : : : : : : : : : : : : : :	
Dd	122	WTIILSNPQANNEIEIRKRLTSAVTKRPFQFGVPYVALSNDSTOKRIAKLESLEFAGSTRPA	181
		: : : : : : : : : : : : : : : : :	
Oy	630	AESQENAFRRKOPAFITQOVLVNIAISLFSGYL	659

CC a sequence present in *Lactobacillus rhamnosus* strain HN001 and encoding a
 CC polypeptide capable of modifying the flavour, aroma, texture, nutritional
 CC and health benefits of milk-derived products, and/or survivability of
 CC microbes in dairy manufacturing processes. The polynucleotides are useful
 CC for improving the properties of microbes used in the manufacture of milk-
 CC derived products such as cheeses, yogurt, fermented milk products, sour
 CC milks and buttermilk; in modifying the flavour, aroma, texture and health
 CC -related benefits of milk-derived products and in increasing the survival
 CC of microbes during industrial fermentation processes. The bacteria may be
 CC used to increase resistance to enteric pathogens and anti-infection
 CC activity, including treatment of rotavirus infection and infantile
 CC diarrhoea, aid in lactose digestion, as anti-cancer and anti-mutagenesis;
 CC liver cancer reduction; reduction of small bowel bacterial overgrowth;
 CC immune system modulation and treatment of autoimmune disorders and
 CC allergies; treatment of allergic responses to foods; reduction of blood
 CC lipids and prevention of heart disease; antihypertensive effect;
 CC prevention and treatment of urogenital infections, *Helicobacter pylori*,
 CC or hepatic encephalopathy; treatment of inflammatory bowel disorder and
 CC irritable bowel syndrome; modulation of endocarditis; and for improved
 CC protein and carbohydrate utilization and conversion. The transgenic
 CC microbial population can be administered to a mammal as an anti-
 CC carcinogenic agent. The present sequence is *Lactobacillus rhamnosus*
 CC outer membrane protein rompa.

XX Sequence 1463 AA;

Query Match 7.8%; Score 251; DB 23; Length 1463;
 Best Local Similarity 25.2%; Pred. No. 1.5e-06;

Matches 165; Conservative 82; Mismatches 295; Indels 114; Gaps 26;

QY 4 HHHHMEGPEVSSNNOSSNMPTIINGQIASNETKSTKASEASPSASSSVSSMFLSAAK 63
 DB 751 HADNMEI--KSLSDAEKQSQI--ALAAASKAAASSAAASAIIVASSASE---ASSA 802
 QY 64 NALISRDALINKNSPTPTSLSEASTSTYTRVAAD---YDPAKSNFTAKGLEN 120
 DB 803 AAASVAAADSANSAASAAAYSYAS--EASAA--AANDSGCYAPASPAASAAAMSA 855
 QY 121 AKTAEYETKMA--DLMAA--LQDMERLANSPDSNHTTEVNNIKKALEAKOTIDIKLK 176
 DB 856 ALSTAGVAAKAVASDAAMAAASAAVAAASQDSKKNQA-----TAATRSQALDLNKK 908
 QY 177 LVTLOQNMKSLTEVLEKTTSDADQIPAINSOLEINKNSADQIIKLERONISYEAVLTNMG 236
 DB 909 I-----KSLTDYASGASS-----SASEAGQAST-----ATSAVAYS 938
 QY 237 EYTKASSEGKIKGLQALQSTVDA---GPOSQAAYLQAOQNNSPDIATKELIDAEIK 292
 DB 939 AASSSASEAGSYAHQAGSSASDAGVSGSAQHAATAASASSYPRDSGIOTSLASQASBP 998
 QY 293 VNELKQEHGLTDSPLVKKAEQISOAKDIOEIKPSGSDPIVGPSSAASGSAAGAL 352
 DB 999 AAKASSNAASATSAAVG-----FSAASDASQOAKTAASADYVA--SSAASTANSASAA 1051
 QY 353 KSSNNSGRISLLIDVDNEMALALQGFPSMTEQFVNNNPATAKELQAMEAQLTAMSDQL 412
 DB 1052 ASATKAG-----DSKAAA-----GFSSAASA--AASSAKGAFAVASEAASAAASDD 1096
 QY 413 VGADGELPAEIQAIKALQALQKOPASDGLATAMGOVAPAAKVGGS-----AGTAGY 467
 DB 1097 VASSA---ASAAAGFPAKASA--AEGAASSAASAAASAAAGTGGASSASEAGAST-- 1151
 QY 468 OMNVKOLYTAFTSTSSSSSYAALSDGYAYKTLINLYSESRGVSATISQ--TANPALS 525
 DB 1152 ---ATSVVAASSASSSASEAGSYAHQAGSSA-----SEATGHASATISQASAAASAS 1200
 QY 526 RSVSRGIESQ-----GRSADASQRAAETIVRDSQTLGDVYSRLOVLSLMST--IYVNPQ 579
 DB 1201 RVPDSGIQSDVSIASSASTASTASASAAQSEASTASTASASHASEQASTASSSDVYSSA 1260
 QY 580 ANQELMQLTASISAKAPQGPYAVONNSADSLQKFAQLEREVDGERSLAEQEN 635
 DB 1261 ASVASSAASAAASAAKAGNSAAGTYSHAASAAASAAKAGSAESQASSAASAAASDD 1316

RESULT 20
 AAG82935
 ID AAG82935 standard; Protein; 5024 AA.

XX AAG82935;
 AC AAG82935;
 XX
 DT 03-SEP-2001 (first entry)
 XX

DE S. epidermidis open reading frame protein sequence SEQ ID NO:2964.

XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis.
 XX

OS Staphylococcus epidermidis.

PN WO200134809-A2.

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30782.

PR 09-NOV-1999; 99QS-0164258.

PA (GLAXO) GLAXO GROUP LTD.

PI Kimmery W;

PR WPI: 2001-316495/33.
 N-PSDB: AAH53785.

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
 DR useful for vaccinating against infections, e.g. endocarditis -
 XX

PS Claim 18; Page 779-781; 2168bp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 5024 AA;

Query Match 7.8%; Score 249.5; DB 22; Length 5024;
 Best Local Similarity 19.8%; Pred. No. 9.1e-06;

Matches 148; Conservative 130; Mismatches 247; Indels 221; Gaps 33;

QY 1 MHHHHHMEGPEVSSNNOSSNMPTIINGQIASNETKSTKASEASPSASSSVSSMFLS 60
 DB 1042 VNNAQHVHDETQATLSSD--TINQLANAVTQAKSNLHGDTKLQHDKDSAKQITAOQLN 1099
 QY 61 SAKNALISRDALINKNSPTD---SLSOLEASTSTSTYTRVAARD----- 103
 DB 1100 SAOK---IHEDSLIDNESRTVOYQHDUTEAQALDGMGLAKESIKDNTNTIYVNGNTIANE 1156
 QY 104 -----YDEA-----KSNFTAKSGLENAKTTLAEYETKMADLMAALOD 140

```
Db 1157 PSKQAYDAAYQNAQNIINGTNOPTINKGVNTTATQTVAKTRKD-----ALDG 1203
Oy 141 MRLANSPDSNNHTEEVNNIKKALEAQKD-----TIDKLNK-LVTIONOKSLTE 189
Db 1204 DHRLEAKKNANOTIRNLNMLNNAQKDAEKLVNSASTLSQVOONLOTAOOLNANME 1261
Oy 190 VLKTTDSADQIPA-----INSOLEINKNSADQIITKDERONISYEAULTNAGEVIRASSDA 245
Db 1262 LRQSTANKQDVAKADSKYLNEDPQIKONIDAV-----ORVETIINETONPELTKAN--- 1312
Oy 246 GIKLQALQSTIVDAGDSQAAYLQAQONNSPDNIATK-----ELIDAEFTVNE----- 295
Db 1313 ---IDQATQSVQNA-----EQALHGAEKLNQDKQTSSTELDGLDLTDQAREKLRQINTS 1365
Oy 296 -----LQKQ-----EHTGLTDSPLYKKAEEQISQAKDIOETKPSGSDIPYVPGSQAASAGS 347
Db 1366 NSRDDIKQKIEQAKALNDA--MKKLEQV--AQKDGVB---ANSDY-----TNEDSQAKDA 1414
Oy 348 AAGALKSS-----NSGRISLLDDVDNEMAALA-----LQGRSMTEQFNVPNNPATAKEL 398
Db 1415 YNNALKQAEADITNNSSNPLNADITNLNLIKQADNLHGAKLQDQKNTTNOALIGNLN 1474
Oy 399 QAMEAQLTAMSDQLVGADGELPAETQALIKDALQALQKPSADGLATAMGOVAFMAAKYVG 458
Db 1475 HLNQPKDALIQAINQATS-----RDQVAKLKE--AEALDEAM----- 1511
Oy 459 GSAGTAGTQVQNVNKKYKTAFASTSSSYAALSDQSAKYKTLNLSYSGRGVQSAISQ 518
Db 1512 -----KOLEDOVODDQISNSSPFINESDQKT---YNKIQAAKEIINO 1554
Oy 519 TANPALRSVSRSGLIESQGRSADASQRAEET--VRD-----SOTLGDVYSRLQ 565
Db 1555 TSNPFLDK-----OKIADTQIKDAVNNLHGDKLQKLAQSKODANNOJLN 1597
Oy 566 VLDSI-----MSTIVSNPOANQ-----EIMQKLTASISKAPQFGYPAYONSADSLQK 613
Db 1598 HLDLITEQKHNFKPLINNADTRDEVNKOLEIAKQUNGDMSTLHK-----VINDKQIQIH 1652
Oy 614 FAQOLERFVQDEGRSLAESQENAFK 639
Db 1653 LS-----NYINADNDKKQNYDNAIKE 1673

RESULT 21
ABP38314
ID ABP38314 standard; Protein, 10182 AA.
AC
XX ABP38314;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3159.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX
XX antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
XX
XX 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI, 2002-381255/41.
```

```
DR N-PSDB; ABN90859.
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
PS Disclosure; SEQ ID 3159; 267pp; English.
XX
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
XX
SQ Sequence 10182 AA;
```

Query Match 7.8%; Score 248.5; DB 23; Length 10182;
Best Local Similarity 19.7%; Pred. No. 2.7e-05;
Matches 141; Conservative 128; Mismatches 256; Indels 189; Gaps 31;

```
Oy 18 NQSSMNPITINGQIASNSERKSTKASEASPSASVSSVSWFSLSAKNLISRLDAIILKN 77
Db 6003 SSDTINQNLNNAVTOAKSNLHGDTKLQHDKDSAKOTIQAOLNLSAQK---IHMEDSLIDNE 6059
Oy 78 SSPTD---SLQLEASTSTSTVTVAADKDEAKSN-----FDTAKSGLENA 121
Db 6060 SRTQVQHDLTBAQLDLGMLGALKESIKDYTNIVSNGYINAEPKQAYDAAYQNAQNI 6119
Oy 122 KTLAEYET--KAMDIAALQDMERLANS-----DPSNNHTEEVNNIKKALEAQKD--- 169
Db 6120 INGTNOPTINKGVNTTATQTVAKTRKDALDGDHRLBEAKKNANQITRNLSNNAQKDAEK 6179
Oy 170 -----TIDKLNK-LVTIONOKSLTEVKTDSADQIPA-----INSOLEINKNSADQI 217
Db 6180 NLVNSASTILEQVQNLQTAQOOLDNANMGLRQSIADKQVAKDSKILNEPQIKONYDDAV 6239
Oy 218 IMLERONISYEAVLTNAGEVIRKASSEAGIKLQALQSTIVDAGDSQAAYLQAQONNSPD 277
Db 6240 -----QVETIINETQNELKAN-----IDQATQSVQNA-----EQALHGAEKLNQK 6283
Oy 278 NTAATK-----ELIDAEFTVNE-----LQKQ-----EHTGLTDSPLYKKAEEQISQA 319
Db 6284 QTSSTELDGLTDLDAQREKLRQINTSNSRDDIKQIEQAKALNDA--MKKLEQV--A 6339
Oy 320 OKDIOEIKPSGSDIPYVPGSQAASAGSAGALKSS-----NSGRISLLDDVDNEMAAL 375
Db 6340 QKDGVBH---ANSDY-----TNEDSAQKDAYNNALKQAEIITNNSSNPLNADITNALNI 6392
Oy 376 A-----LQGRSMIEQFNVPNNPATAKELOAMEAQLTAMSDQLVGADGELPAETQALIKDAL 430
Db 6393 KQADONLHGAKQKLOQDKNTTNOALIGNLNHNGPQKDALIQAINQATS-----RDQV 6443
Oy 431 AQALQKPSADGLATAMGOVAFMAAKYVGGSACTAGTQVQNVNKKYKTAFASTSSSYAANA 490
Db 6444 AEKLEKE--AEALDEAM-----KQLEDOVNODDQISNSSPF 6476
Oy 491 LSDGSAYVKTLSLSYSESQVQSAISOTANPALRSVSRGIESQGRSADASQRAEET 550
Db 6477 INEDSDQKPT---YNDKIQAAKEIINQTSNPTLTK-----OKIADTUL 6515
Oy 551 ---VRD-----SOTLGDVYSRLQVLDL-----MSTIVSNPOANQ-----EI 585
Db 6516 QNIKDAVNNLHGDKLQKLAQSKODANNOJNLHLDLITEQKHNFKPLINNADTRDEVNKOLEI 6575
Oy 586 MOKLTASISKAPQFGYPAYONSADSLQKFAQLERFVQDEGRSLAESQENAFK 639
Db 6576 AKQUNGDMSTLHK-----VINDKQIQIHL-----NYINADNDKKQNYDNAIKE 6619
```

RESULT 22
AAU37018 standard; protein; 3158 AA.
ID AAU37018;
XX AAU37018;
AC AAU37018;
XX
DT 14-FEB-2002 (first entry)
XX
XX Staphylococcus aureus cellular proliferation protein #1188.
DE
XX
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
OS
XX
XX WO200170955-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001WO-US09180.
PF
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Haselbeck R, Ohlsen KL, Zyskind JM, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
PI
XX
XX WPI; 2001-611495/70.
DR N-PSDB; AAS54877.
XX
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
XX Example 3; Seq ID No 12611; 511pp; English.
PS
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 3158 AA:
Query Match 7.7%; Score 247; DB 22; Length 3158;
Best Local Similarity 20.6%; Pred. No. 7e-06;
Matches 167; Conservative 134; Mismatches 285; Indels 226; Gaps 34;
QY 10 SGPESVSSNGSSNMPITNGIASSEKRETKASEASPSASSSVSSMSFLSAKN----- 64
DB 43 TGIPDVVYVPSEITAIANKYTKAOKNDLNGNTNLATKAKONVHAIDLPNLNOQRDEYSK 102

QY 65 -----ALISLRDAILUNKSSPTDLSQLEASTSTSTVYRVAANDYDEAKSNFPTAKSGLE 119
DB 103 QITQATLVPVNAIIQQAATTI LNDAMTQKGIANKKAQIK -GSENYHDADPDKQTA---YD 158
QY 120 NAKTIAEYETKMA-----DILMAALQDMERLANDPSNNHTEEVNNIKKALEAKQRTIDKL 174
DB 159 NAVTKAEELKQTPTNPMDPNTIOQALTQVNDPNOALNGNOKLADAKQAKTMLGTLDHL 218
QY 175 N-----KLVILOL--ON-KSLTEVLKTTDSA--DOLPAINSQLEINKNSA 214
DB 219 NDAQOKALTTQVQADPIATVNNVKNQANQNLNANLNALNADKTETLNS---INFTDA 275
QY 215 DQIITKLERONISY-EAVLTNAGEVTKASSEAGIKLQALQSTVDA-----GDOSQAAV 267
DB 276 DQAKKDAVTVNAVAHAEGLISKA-----NGSNAQTEVEQAMQRYNEAKQALNGNDVYQRAK 311
QY 268 LQAQO--NSPD-NIA---ATKELIDAEI--KYNLEKQPHLTGTLSPLYKKAEEQISQA 319
DB 332 DAARQVTTNANDLNOAQKDALQOVDAAQTVANVNTIKQTAODL-----NQAMTQLKQG 385
QY 320 QKDIOETKPSGSDIPIVGPESGSAASAGSA-----AGALKSSNNSGRISILLDDV----- 368
DB 386 IADKDDTKANGNFVNAADTKQNAVNNNAVAHAEQIISGTPMANVDPQVQAALQVYQAKG 445
QY 369 ---DNEMAAIQAQFSSMIEQF--NVNNP-ATAKELAMEAQTLA-----MSDQLVGAD 416
DB 446 DLNGNHNLIQAKKNANMTAIDQLPNLNQPRTALKDQVSHAEIVTGVNAIKONADALNNAM 505
QY 417 GELPAELQA----- 425
DB 506 GTLKOQIOANSQVPOSVDPTQADODKQOAVYNNANAQOITANGTPTPVTLPTDTVTAVTT 565
QY 426 ---IKDA-----LQPSADGLATAMGOV-AFAAAKVGGSAG 462
DB 566 MNOAKDALNDEKLAQAKQDALANLDTLRDLNQPORDALRNQNOQAALATVQOTQONSQ 625
QY 463 TACTVQNNVQKLYKTAFFSSSTSSSYAAALSDGSAVKTJNSLYSEKSGVQSAISQTPAM 522
DB 626 NVNTAMGNLEKQGIANKDTYKASENYHDADKQTA-----YTNVSAQEGIIQNTNP 678
QY 523 AL-----SRSV-----SRGIESQGSADASQA-----AETIVRSQTLGDVYSRLQVL 567
DB 679 TLNPDEITRALTQVTDKANKLNGEAKLATEKQNAKDAVNMTILNDAQ----- 726
QY 568 DSLMTSTVSNPQANQOEIMQKLPAISAKAPQFG-YPAVQNSADSLQKFAQLRPREFYDGE 626
DB 727 -----KALAGQIDQSPETATVTVQKQTAISLDQAMNQLSQALINDKT 768
QY 627 RSLAE-----SQENAFRRQPAFTQOVLVN 650
DB 769 QTLTDGNYLNADDPDKQNAV-KQAVERAFAALLN 799
RESULT 23
AAU75489 standard; protein; 2659 AA.
ID AAU75489;
XX AAU75489;
AC AAU75489;
XX
XX 23-APR-2002 (first entry)
DT
XX
XX
XX S. aureus antigenic protein associated protein #9.
DE
XX
XX Antigenic protein; vaccine; SEPRX; antibacterial;
KW antiinflammatory; dermatological; antituber; tuberculosatic;
KW immunosuppressive; septicaemia; food poisoning; skin disorders;
KW peritonitis; endocarditis; tuberculosits; blood infection; sepsis;
KW meningitis; pneumonia; stomach ulcer; gonorrhoea; necrotising fasciitis;
KW impetigo; Lyme's disease; gastro-enteritis; dysentery; shigellosis.
OS Staphylococcus aureus.
XX
XX WO200198499-A1.
PN

XX 27-DEC-2001.
PD
XX
XX 20-JUN-2001; 2001WO-GB02685.
PF
XX
XX 20-JUN-2000; 2000GB-0014907.
PR
XX
XX (UYSH-) UNIV SHEFFIELD.
PA
XX (BIOS-) BIOSYNEXUS INC.
PA
XX Foster S, McDowell P, Brummell K, Clarke S;
PI
XX WPI; 2002-106544/14.
DR
XX
XX Identifying antigenic polypeptides expressed by pathogenic organisms
PT e.g., *Staphylococcus aureus* during infection, by SEREX (serological
PT identification of antigens by recombinant expression cloning)
PT techniques -
PS
XX Disclosure; Page 64-72; 85pp; English.
XX
XX The invention relates to a method for identifying antigenic polypeptides
CC expressed by pathogenic organisms e.g., *Staphylococcus aureus* during
CC infection, by SEREX (serological identification of antigens by
CC recombinant expression cloning) techniques. The method involves
CC providing a nucleic acid library encoding genes/partial gene
CC sequences of pathogenic organisms, transforming/transfecting
CC the library into host cells, contacting the polypeptides expressed by
CC the genes with autologous antisera derived from an animal infected
CC with, or has been infected with the pathogen and purifying the nucleic
CC acid encoding the polypeptide or partial polypeptide binding to the
CC antisera. Also included are the nucleic acids and polypeptides
CC isolated by the method, vectors and transformed cells expressing them, a
CC vaccine comprising the polypeptide and the production of monoclonal
CC antibodies against the polypeptides. The protein and vaccine are useful
CC for immunising an animal (preferably human) against a pathogenic microbe.
CC The proteins and antibodies are useful for manufacturing a
CC medicament for treating *Staphylococcus aureus*-associated septicaemia,
CC food poisoning or skin disorders or *Staphylococcus*
CC epidermidis-associated septicaemia, peritonitis, endocarditis,
CC tuberculosis, blood infections, sepsis, meningitis, pneumonia, stomach
CC ulcers, gonorrhea, necrotising fasciitis, impetigo, Lyme's disease,
CC gastro-enteritis, dysentery and shigellosis. The present sequence
CC is an *S. aureus* protein sequence associated with the antigenic
CC proteins of the invention.
CC Note: The present sequence is included in the sequence listing but
CC is not mentioned anywhere else in the specification.
CC
XX
XX Sequence 2659 AA;
SQ
Query Match 7.6%; Score 244.5; DB 23; Length 2659;
Best Local Similarly 20.1%; Pred. No. 7.9e-06;
Matches 157; Conservative 137; Mismatches 285; Indels 203; Gaps 34;
OY 7 HMEGSPESVSSNMPIINGOLASNETKESTKASEASASSSSWSFSLSAKAL 66
DB 1779 HQDAVNTIKONATNLSAM-GNLRQAVADKQVKTEEDYADADTA-----KONATNSA 1831
OY 67 ISLDAALINKSPDLSLEASTSTSTYTVRAAKDYDE-AKSNFTYA----- 114
DB 1832 VSSAETIINQNTPTMTSDVDVNRATSAVTSNMKNALNGEKAQSKTDAARAIDALPHLNN 1891
OY 115 -----KSGLENAKTLA-----EYETKADLMAALODMRLANSDPSNNHTTEVN 158
DB 1892 AOKADVASKINMAASNIAGVNTVKKQGTDLNTPAMGNLQCAINDEQTTLN---SQNYODATP 1948
OY 159 NIK-----KALEAOKDTJDKLKLVTLQNNKSLTEVLKTTPSAOQDAINSOLEIN---- 210
DB 1949 SKRTATYNAVAOAAKDIILKSN-----GOKTKPOV---TEAMQVNSAKNNLGGTRILD 1999
OY 211 --KNSADOIIRKDERONISYEAVLTNAGEVITKASSEAGIK-----LGOAL-----OSI 256
DB 2000 OAKOTAKOQLNNMTHLTAAQKTNLTN--QINSCTTAVAGVGVQVGSNANTLDQAMNTLRQSI 2057

OY 257 -----VDAGPOSQA-----VLQAOON----- 273
DB 2058 ANKDATKASEDYVDANNDKOTAYNNAAAEFTIINANSNPEMNPSTITOKAEQVNSKTA 2117
OY 274 -NSPDNIAATKELIDAEATKYNELKOEHTGLTDS-----PLVKRAE 314
DB 2118 LINGDENILAAKQ---NAKTYLNTL---TSTIDAKQNNLSQITSATRVSGVDIVKQNAQ 2170
OY 315 OISOAKDIO-----EIKPS-----GSDIPIVGPSSAASAGSAGALSSNNSGR--I 361
DB 2171 HLDQAMASLQNGINNESQVSSSEKRYRADPTKKQDEYNALITAAKAILNKSTGPPTAQNAV 2230
OY 362 SLILDDVDNEMAAIALQFRSMIEQFNVPNNPATAKELQAMEQOLTAMSDDLQVADGELPA 421
DB 2231 EAALQRYNN--AKALNGDAKLIAQNAKOHGLCTLHTHTTAQNDLITNOI----- 2279
OY 422 EIQAIKDALQALQKPSADGLATAMGOVAFMAAKVGGSA-----G 462
DB 2280 -SQATNLAGVESVKQ--NANSLDAGMNLQTAINDKSGTLASQNLDADEQRNAYNOAVS 2337
OY 463 TAGIVQANVQKLYRTAASSTSSSYAAALSDGYSAYTTLNLSYSESSGVOAISQIANP 522
DB 2338 AAEFTI-----LNKQTPNPAKTAVEQALNNVNNAKHALNGT--ONLNNAKQAITTAINGA 2390
OY 523 ALSRSVRSRGIESQGRSADASORA--AETIVRDSQTLGDVYSRLQ--VLDSLMTIVSNPQ 579
DB 2391 SDLNQKQKDALKAQ---ANGAQRYVSNADQVQHNATELMTANGTLKHAIDAKTNLASSKY 2447
OY 580 ANDEIMQK-LTASISRAPOF--GYPAVONSADSLQFAQLE--REFVDSERSLAESOE 634
DB 2448 VMASTQKQNTATVTVAHHEHISTPIVYTPPSEVTAANQVNSAKQELNDEKRLREAKQ 2507
OY 635 NA 636
DB 2508 NA 2509
RESULT 24
AAU034339
ID AAU034339 standard; Protein: 2434 AA.
XX
XX AAU034339;
AC
XX 14-FEB-2002 (first entry)
DE
XX *Staphylococcus aureus* cellular proliferation protein #615.
XX
XX Antisense: prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
KW
XX *Staphylococcus aureus*.
XX
XX WO200170955-A2.
PD 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.

DR N-PSDB: AAS52198.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3: Seq ID No 5835; 511bp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2434 AA;

Query Match 7.5%; Score 239.5; DB 22; Length 2434;
Best Local Similarity 21.4%; Pred. No. 1.4e-05;
Matches 147; Conservative 107; Mismatches 222; Indels 211; Gaps 31;

QY 21 SMNPITNGOIASNETSEKSTKASEASPSASSSVSSWFLSKAKNALISRDATLNKNSP 80
DB 1889 AMERLING-IQKDYKQKQSVNFTDADPEK-----QTAVNNAVTAENITINQANGT 1937
QY 81 TDSLQLEASTSTSTVTRVAA---KDYDEAKSNFTAKSGLEN----- 120
DB 1938 NANOQVEAALSTVTYTKQALNGDKRVTDKANNANQTLSTLDLINAQGVATGNINQAH 1997
QY 121 -----AKTAEYETKNAADLMAALQD-----MERLANSNP--SNHTEEVNNTKKAL 164
DB 1998 TVAEVTQAIQTAEQELNTAGNLKNSLNDKDTLGSQNFADADPEKKNAVNAEHNAENLT 2057
QY 165 EAQKDTIDKLKLVTLQONKNSL-TEVLKTPDSADQIPAINSOLEINLK---NSADQIIND 220
DB 2058 -----NKSTGTNPK-----DOVEAAMQVNAKTKALNCTQMLEKA 2093
QY 221 LERONISYEAV--LTVAGEVIRKASSEAGIKGQALQSIYDAGDQSGAAVLAQOQNSPN 278
DB 2094 KOHANTAIIGLSHETNAOK-----EALKQV-----QGSTVAAEQNGQOKAN 2136
QY 279 IATKELIDAAETKV-----NELKQEHGTLDSPVYK--AEQISQAKQIDQETKPSG 330
DB 2137 -----NVDAADKRLQSIADNATTKQNONYTDASQNKDAYNNAVTTQGIIDQTTSP 2190
QY 331 SDPIVPGSGAASAGSAGALKSSNNSGRISLLDDVNNEMAIALGFRSMIEGFVN 390
DB 2191 LDPTVY-----NOAAQGVSTTKNA-----LNGNEN-LEAAKQQAQSSQISGLDNN 2234
QY 391 NPATAKELQAMENQJLAMSQDLVGADGELPAETQATKDALAQAQKPSADGLATANGOVA 450
DB 2235 N-----AQKQTVTDQINGA-----HWYDE--ANQIKQ-NMQNLNTANG---- 2269
QY 451 FFAAAKVGSGAGTAGTGVQNNKQLYKTAFSTSSSSYAAALSDGYSAYVTLNLSSESS 510
DB 2270 -----NLKQ-----AIDKDKATKATVNF-----T 2288
QY 511 GVQSAISQTFANPALSRVSRSIGESGRSADASQRAAETIVD-----SQTLGDVYSRLQV 566
DB 2289 DADQAKQQAQYNTAVTVA---ENITISKANGNATQAEVEQATIKOVNAAKALNGNANVQHA 2345

QY 567 LDSLMSTIVSN---PQANOEIMQKLTASISKAPQFGYPAYONSADSLQKFAQJEREFV 623
DB 2346 KDEATRLINSSNDLNOAQKDALQOVNAVTVAA---GVNNVKTQAEQLNNMATTQKQIA 2402
QY 624 DGERSTAE-----SQENAFRKQPA 642
DB 2403 DKEQTKADGNEFVNADPDKONAVNAQAVA 2429

RESULT 25
ABP39618
ID ABP39618 standard; Protein; 2137 AA.
XX
XX ABP39618;
AC
XX
XX 24-JUL-2002 (first entry)
DT
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4463.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
KW
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2002-381255/41.
DR N-PSDB; ABN92163.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
PS
PS Disclosure; SEQ ID 4463; 267pp; English.
XX
ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 2137 AA;

Query Match 7.4%; Score 237; DB 23; Length 2137;
Best Local Similarity 19.9%; Pred. No. 1.7e-05;
Matches 136; Conservative 134; Mismatches 316; Indels 96; Gaps 23;

QY 9 ESGPESVSSNOSMNPITNGOIASNSE-----TKSTKASEA---SPSASSVSSWFLS 60
DB 1079 ESDSTSTSLSESTSTSLGSGTASSTSDASTSTSESDTSESTSLSESLSTVSOSTSAS 1138
QY 61 SAKNALISLRDAIILKNSSPTDLSQLEASTSTSTVTRVAAADYDEAKSNFTAKSGLEN 120
DB 1139 TSESASTSTSES--ESNASTSLSGSLSTISDSTSTIS---DSASTS--TSESESDS 1190
QY 121 AKTLAEYETKNAADLMAALQDMERLANSNPSSNNHTEEVNNTKKALEAQKDTIDKLKLVTL 180


```
Db 302 NNLISQITSAITRVSGVDIVKQNAOHLDOAMASLQNGINNESQVKSSEKRYROADTKKQOEY 361
Qy 158 NNIKKALEAOKDTIDKLKLVTLONOKSLTEVLKTTSDAOIPAINSOLEINKKSADQI 217
Db 362 DN---AITAAKAILLKSTGPTAQN---AVEAALQRVNNAKDALNGDAKLIAOQNAAKOH 415
Qy 218 IKDERQNISEAVLVTNAGEVIKASSEAGIK-----LGO-----AL 253
Db 416 LGTLHTTTAQRNDLTN--QISQATNLGAVESVKONANSLDGMGNLOTAINDKSGTLAS 473
Qy 254 OSIYDAGDOQAAYLQA-----QQNNSPDNIATKELIDAEETKYNELKOEHTGLTD 305
Db 474 QNFLDADQKRAVYNAQVSAEITLTKOTGPN---TAKTAYEQALNNVNNNAKHALNGQN 530
Qy 306 SPLVYKA-----EEQISOAKDIOEIKPSGS-----DIPYIPGSGSAASAGSAGA 351
Db 531 LNNAKQAALITAINGASDINOKOKDALKAQANGAQRVSNADV-----OHNAETELNTAMGT 585
Qy 352 LKSSNNSGRISLLDDVDNEMAALOGFRSMIEQFVNNNPATAKEL-----QAMEAO 404
Db 586 LKHA-----IADKTNTLASSKYVNAADSTKQNAVTTKYTNAEHITSGPTVYTTTSE 636
Qy 405 LTAMSDOLVGADGELPAE--IOAIK-----DALAQALKOPSDAGLATAMQOVAFAAAK 455
Db 637 VTAANQVNSAKOELNGEBRLKAKONANTALDALTO--LNPQKAKKEQVGO-----688
Qy 456 VGGSGAGTAGYQAM-----VKOLYTAHSST---SSSYTAALSDGYSAIKYKLNILY 505
Db 689 --ANLEDDYQVYOTNGOALNNNAKGLRDSIANETTVKTSONTDASPNNOSTY---NSAV 743
Qy 506 SESRSGVQSAISOTANPALSRVSRSIGESOGRSADASORAETIVRSOTGLGDVYSRLQ 565
Db 744 SNAK-----GIINQTNPTMTDSAITQA--TTOYNNAKNGLNGENIRMANONTAKOMLNTLS 798
Qy 566 VLDSIMSTVSNPQANOEIMOKLTASISKAPQFGYPAVQNSADSLQKFAQLER-----620
Db 799 HLTNNOKSAISS-QIDRAGHVSEVTAT-----KNATETELNTOMGNLEQAIHQ 845
Qy 621 -----EFVGERSLAESQENAFKOPAFIOQ 646
Db 846 NTVKOSVKTADKAKKRDAYTNAAVSRAEAILNK 878

RESULT 31
AAU37490
ID AAU37490 standard; Protein: 1048 AA.
XX
AC AAU37490;
XX
Df 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #1660.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN MO200170955-A2.
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
```

```
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI: 2001-611495/70.
DR N-PSDB: AAS55349.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 13083; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequences is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 1048 AA:
XX
Query Match 7.3%; Score 234.5; DB 22; Length 1048;
Best Local Similarity 20.5%; Pred. No. 9.3e-06;
Matches 154; Conservative 132; Mismatches 288; Indels 179; Gaps 32;

Qy 10 SGPEVSSNOSMNPITINGQIASNETKESTRKASASVSASSVSMFSLSAKALLSL 69
Db 201 AGVOTVQSNATLDDAMN--TLRQSIANKDARKASEDYVDVAND-----KOTAYNNAAVA 253
Qy 70 RDAIILNKNSP-----TDSLQLEASTS-----TSTVTRVAAKDY-----DEAK 108
Db 254 AETIINANSNPEBNPSTTQAKAEQVNSSKTLANGDENITAAKQAKKTYLNTLTSTDQK 313
Qy 109 SNF-----DTAKSGLENAKTLAEY-ETRMADLMAALQDMERLANS-----DPSNNHTEEV 157
Db 314 NNLISQITSAITRVSGVDIVKQNAOHLDOAMASLQNGINNESQVKSSEKRYROADTKKQOEY 373
Qy 158 NNIKKALEAOKDTIDKLKLVTLONOKSLTEVLKTTSDAOIPAINSOLEINKKSADQI 217
Db 374 DN---AITAAKAILLKSTGPTAQN---AVEAALQRVNNAKDALNGDAKLIAOQNAAKOH 427
Qy 218 IKDERQNISEAVLVTNAGEVIKASSEAGIK-----LGO-----AL 253
Db 428 LGTLHTTTAQRNDLTN--QISQATNLGAVESVKONANSLDGMGNLOTAINDKSGTLAS 485
Qy 254 OSIYDAGDOQAAYLQA-----QQNNSPDNIATKELIDAEETKYNELKOEHTGLTD 305
Db 486 QNFLDADQKRAVYNAQVSAEITLTKOTGPN---TAKTAYEQALNNVNNNAKHALNGQN 542
Qy 306 SPLVYKA-----EEQISOAKDIOEIKPSGS-----DIPYIPGSGSAASAGSAGA 351
Db 543 LNNAKQAALITAINGASDINOKOKDALKAQANGAQRVSNADV-----OHNAETELNTAMGT 597
Qy 352 LKSSNNSGRISLLDDVDNEMAALOGFRSMIEQFVNNNPATAKEL-----QAMEAO 404
Db 598 LKHA-----IADKTNTLASSKYVNAADSTKQNAVTTKYTNAEHITSGPTVYTTTSE 648
Qy 405 LTAMSDOLVGADGELPAE--IOAIK-----DALAQALKOPSDAGLATAMQOVAFAAAK 455
```

Db 649 VTAAANOVNSAKOELNGEDERLEREKQVANAANAIDALITQ-LNTPOKAKIKEVQC----- 700
Qy 456 VGGSGAGACTGVOMN-----VKQLYKTFASST-SSSSVAALSDGSAAYKTLSLY 505
 ||| : | : | : | : | : | : | :
Db 701 -AARLEDEVQVVGNGCALNNAMKGLDJSIANETTVKTSQNTVDASPNNQSTV---NSAV 755
Qy 506 SESMSGVSALSICQANRALSRSVRSRGSIESQGRSADASQAETAETIVRDSOTLGDVYSRLQ 565
 : : : | : | : | : | : | : | :
 : : : | : | : | : | : | : | :
Db 756 SNAR---GIINONNNPMTDTSALTQQ-TQOVNNAKKKLGAENLRMAONTAKQNLTLS 810
 : : : | : | : | : | : | : | :
Qy 566 VLDSLMSTRIVENPONOEIMOKLTATISKAPOGCVPAVONSADSLCKFAALER----- 620
 | : : : | : | : | : | : | : | :
Db 811 HLTNNQKSATIS-QIDRAGHYSEVTAT-----KNAATELNTQMGNLQAIHQ 857
 : | : : : | : | : | : | : | :
Qy 621 -----EFYDERSLAESQENAFRRKPAPIQQ 646
 : | : : : | : | : | : | : | :
Db 858 NTVKOSYVFETDADKAKRDATYNNAVSRRAELLINK 890
 : | : : : | : | : | : | : | :

RESULT 32
AAU37464
ID AAU37464 standard; Protein: 560 AA.
XX
AC AAU37464:
XX
DT 14-FEB-2002 (first entry)
XX
De Staphylococcus aureus cellular proliferation protein #1634.
XX
KW Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
Pn WO200170955-A2.
XX
PD 27-SEP-2001.
PF 21-MAR-2001; 2001WO-US09180.
Pr 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-20684BP.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
Pa (ELIT-) ELITRA PHARM INC.
Pi Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
Pi Yamamoto RT, Xu HH;
XX
DR WPI: 2001-611495/70.
N-PSDB: AANS5323.
New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids -
Example 3; Seq ID No 13057; 51pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antipodes capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen

	CC	for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences .
SQ	Sequence	560 AA:
	Query Match	7.3%; Score 234; DB 22; Length 560;
	Best Local Similarity	20.8%; Pred. No. 4.4e-06;
Matches	127; Conservative	108; Mismatches 239; Indels 138; Gaps 21.
OY	ISSAKNALISLDALINKNSPTDLSOLEASTSTVTVRVAAKDYDEAKSNFDTAKSGL	118
Db	1 MSSSTRSESRSLSDSTSSESGSTSTSLNNTSGCSISTSTSISS-----ESTSF--KKSS	52
OY	119 ENAKTLAEYEETKMAD---LMAALLDMERLANDPSNNHTEEVNNIKKLAEOKDITDKLN	175
Db	53 VSTSLMSMTSTSLSDSTSLSTSLSDTSDDSKD-----SLTSMSTDSI-	97
OY	176 KLVLTIONOKKSILTEVLKTTPDSADOIPAINSOLEIKNKNSDOIHKDLRONIYE-----A	230
Db	98 -----STKSDDSI STSTS--ISGSTEESDSTSSSEKSDSTMSISMOSTSGS	146
OY	231 VLTNAGEVIKASSEAGIKIGQAL-QSYVDAGDOQAVALPDAQONNSPDNIATKELIDAA	289
Db	147 TSTSTSTSLSDSTSTSLSLASMNGGVDSNAQSASNSTSTSTNSESOQTSS-YTSQ	205
OY	290 ETKVMLEKOHTGLDSPLYKAEROI-----SQACKDIQEIKPGSDPIPYGP	338
Db	206 STSGSESTSTSTSLSDSTSIKSTSGSGSVTSAHSLSGESSEDOSISTSAFE-----	259
OY	339 SGSASAGSAACALKSSNNSGRISLLDVDENEMAIALOGRSMIEQPNNNPATAKEL	398
Db	260 -STSESASTSLSDSTSTNSGSAS-----TSTSLNNSASAF-	295
OY	399 QAMEAOQLTAMSPDOLVGADDELPAEIQAIKDALAQALKP SADGLATAMGOVAFAAKYGG	458
Db	296 --SDLSTSLSDSTSGASMGSSSDSQSTYSALSJDSLSTSNRMST----IASLSTSVGT	349
OY	459 GSAGTAGTYQMNVKQLITAFBSSTSSSYTAALLSDGYSAKYKLNLISYSRSGVQSAISO	518
Db	350 SESGS-----TSESTSESDYSTSLSDSQSTSR-----STPASGSASTST	390
OY	519 TANPALRSVSRSGLIESQRSADAQQ-----RAEFIVYDSOTLGDPVYSRLQVLSL---	570
Db	391 YSD-SRRSTASSTSTSMRTSTSDSQSMLSJTSTISMDSYSLSD-----SVSDSTSDS	442
OY	571 -----MSTIVSNPANOEEIMOKLTATISKAPOGFYAVONSADSLQKFAOALER	620
Db	443 TSASITSNGMSVSLSDSTSTSTSTASAEWMSAISDSQMSSEV--NDSSVSSENSSEDS	500
OY	621 EFVDEGRSLAES	632
Db	501 KMSGSTSVSDS	512
RESULT	33	
ID	ABB54168	
AB	ABB54168 standard; Protein; 1063 AA.	
XX	AC	ABB54168;
XX	DT	16-MAY-2002 (first entry)
XX	DE	Lactococcus lactis protein yjhd.
XX	KM	Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX	OS	Lactococcus lactis IL1403.
XX	XX	

```

PN FR2807446-A1.
XX
XX 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
XX WPI; 2002-043418/06.
XX
XX New nucleotide sequence useful in the identification of Lactococcus
XX lactic and related species -
XX
XX Claim 6; SEQ ID NO 870; 2504bp; French.
XX
XX The present invention is related to a Lactococcus lactis nucleotide
XX sequence (ABA90521) and related proteins (AB53300-AB55621). The
XX nucleic acid sequence is useful in the detection and/or amplification of
XX nucleic acid sequence, particularly to identify Lactococcus lactis or
XX related species. The proteins of the invention are useful for the
XX biosynthesis or biodegradation of a composition of interest. The
XX invention helps research in lactic bacteria, particularly useful in the
XX production of yogurt and cheese.
XX
XX Note: The sequence data for this patent is based on equivalent patent
XX WO200177334 (published 18-OCT-2001) which is available in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 1063 AA:
XX
XX Query Match 7.3%, Score 234; DB 23; Length 1063;
XX Best Local Similarity 22.4%; Pred. No. 1e-05;
XX Matches 160; Conservative 108; Mismatches 285; Indels 160; Gaps 29;
XX
OY 13 ESVSNOSSMNP-----INGQINSSEKSTKASEAS--PSASSSVSMFPL-- 59
DB 146 DSSSSNDGNSISLSSNNAADSVSAVGSOSTSSGVLSSESAIDSGIASVSGSSENL 205
OY 60 -----SSAKNALISLRDAILNKNSSPTDLSQ-----LEASTSTSTVTRVAAKD-YDEAK 108
DB 206 VGNSSASASAAVASFMTILATNPSMVPMLTGALAAAPATSSGAILNTITGLDVNQAI 265
OY 109 SNFDTAKGLENA-KTLAEYETK-MADLMALQDMERLANSOPSNHTEEVNNIKALEA 166
DB 266 ST--VGISGLANIFSTLCTFNIPGWTAAALNGVEQIVNI-----VGNIQEA--- 311
OY 167 QKDTIDKLKLVTLQNKSKLFEVAKTTDSADQIPAINSOLEINKNSADQITKDLERONI 226
DB 312 -----AANPGAFLNLEKRS-----AGLDVSQIPLVGQIAAFAAFA-----PSM 350
OY 227 SYEAVLT--NAGEVIKASSEAGIKLGOALQSVDAQDSQAVALQAOONNSPDNI--AAT 282
DB 351 SPAAMLTFLNPTIPGLSSIRASL--VLSPLVLSAISTVTSQIVN-QLNTTISNMLGVN 407
OY 283 KELIDAEFTKYNELKOEHTGLTDSPLVKKAEQISOAKDIOETIKPSGSDIPIVGPSG-- 340
DB 408 FDLDTLVLQGDNDLVNVLGYVNSAIRNV-QQIAMSQ-----LSPTISNIPLVGTTVNN 461
OY 341 -----SAAASGSAAGALKSSNNSGRISILLDDVDNEMAAILQGFERSIIEFPNV-- 389
DB 462 VLSPTLNNLTGASLGEVA-----NLTVGSVSLLDQVNNNSLGNLISLSTALATIENTLQ 514
OY 390 -----NNPATAKE-----IQAMEAQITAMSDQVLGADGELPAEQIAIKDLAOLKQ- 436
DB 515 NSLNSFGNLPAGASDILNQVLNQDLNATNNIYESATGIYNNRPG-LGAIENLSMTIQSI 573
OY 437 -----PSADGLATAMGOVA-----FAAAKVGGSAGACTYVOM 469
DB 574 PNNINNVNNAALNGITITINSLTSPVGASTVNPNSANSOSSSSASSSSASSSSSTSS 633

```

```

OY 470 NVKQLYKT-----AFSSSTSSSYAALSDGYSAKYKTLNLSYESRSGVOSAIQOTANPA 523
DB 634 NVSSNTSSNSBRANTSSSTSNANSSSSSSSESSA-----SSSNSSESSVASSSSVDSSQS 688
OY 524 LRSYRSRGIESQGRSADAPRAETIYRDSOTLGDVYSRLQVLDLSMTIYSNPOAOE 583
DB 689 SSAGVNSSSSSSAEQSSASSSSSESSVASSS-----VDSQS---SSAGVNS 734
OY 584 EIMQKLTASISAKAPQFYPAYONSADSLQKFAQLEREFVDEGRSLASQENA 636
DB 735 SSSSESSASSSSSESSGVASSSSVDSSQSSAGVNSSSSSSAEQSSASSSNS 787

RESULT 34
AAU34207
ID AU34207 standard; Protein: 2025 AA.
XX
XX AU34207;
XX
XX 14-FEB-2002 (first entry)
XX
XX Staphylococcus aureus cellular proliferation protein #483.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX
XX 23-MAY-2000; 2000US-206848P.
XX
XX 26-MAY-2000; 2000US-207727P.
XX
XX 23-OCT-2000; 2000US-242578P.
XX
XX 27-NOV-2000; 2000US-253625P.
XX
XX 22-DEC-2000; 2000US-257931P.
XX
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KI, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX
XX N-PSDB; AAS52066.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 5703; 511bp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at

```

```

CC ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2025 AA:
Query Match 7.3%; Score 233; DB 22; Length 2025;
Best Local Similarity 19.8%; Pred. No. 2.7e-05;
Matches 141; Conservative 129; Mismatches 275; Indels 166; Gaps 25;
QY 10 SGPESVSSNOSNPPIINGIASNETKSTKASEASPSASSSVSWFSLSAKNAIISL 69
DB 1078 NGQONLANAKDKANAFVNSLNGINGNOOQ-----DLAHNAIINADIVSDTIDVINOQIDL 1131
QY 70 RDAIILKNSSPTDSLQSL---EASTSTSTVTRVAANDYDEAKSNFTPASGLENAKTLAE 126
DB 1132 NDAM-----ETLKHLYDNEIIPNAEQFTVNYONAD--DIAKTNFDDAK----- 1170
QY 127 YEKRMADLMAALQDMERLANSNHNTEVEVNNIKKALEOKRTIDKN-----KLV 178
DB 1171 -----RLANFTLLNSDNT--NVNDINGAIGQVDAIGNLNGEORLQEAQK 1213
QY 179 TLQONKSLTEVLTQTTDSA-----DOIPAINSOLEINKNSADQIILDLERQNTSYEAVLT 233
DB 1214 AIONVKVVLADKLEIASNATDQDKLIAKNAEELANSIINNINAKTSNQDVSYQOTAG 1273
QY 234 N-----AGEVIRKASSEAGIKLGQALQSTVDAGDQSOAAVILQAOQNSPDNIATKELI 286
DB 1274 NQAIIEGVHANEIIPKAKIDANKVDKQVQAL-----DEIGRPNPLT 1314
QY 287 D-----AETKRVNE-LKQERTGLTDS---PLVKAEEQISOAQDKQIEKSSGSDIPIVCP 338
DB 1315 DKRKQALKRINQIILQOGHNDINNALTKEAIEQAKERLADODIDLVYAKED----- 1368
QY 339 SGSAAGASAGALAKSSNNSGRISL-----LLDQVDNEMAAIATLOGFRS--M 383
DB 1369 -----AKNKIKALANAKKRDQINSNDLTPEQAKALKKEID-EAEKRALONVEAQT 1418
QY 384 IEQFN-----VNN-----PATAKELQAEQTLTMSDQLVGADG-----ELP 420
DB 1419 IDQLNGNLGLDIDIRNTHTWEVDQPAVNEISEATPEQLLVNGELIVHRDDIITTEDVL 1478
QY 421 AELQATKDALQALQKPS---ADGLATAMGOVAFAAKYGGSSAGTACVQMANVQLYK 476
DB 1479 AHAHNLIDQLAEVYIDPPTATISDSTLAKVEVTLIDGSKV-----IVNVPKRVK 1529
QY 477 --TAFSTSSSYAALSDGYSAYKTLNLSYSESRSQVQSAISQFANPALRSVRSQIE 534
DB 1530 ELSVVAQQAIESEIENAVQKININNSVTLTEQKEAIAIEVKKLQQAIDHNNAPDVH 1589
QY 535 SOGRSADASORAAETTVRDSQTLGDVYSR-LQVLDLSLMTSTIVSNPOANOEIIMOKLTAST 593
DB 1590 SVEEIQOQEQAHIEQFPNPEQFTTEQAKSNAIKSIDAIIQHMIDEIKARTDULTQEKQEA 1649
QY 594 SKAPQGYPAVQ--NSADSLQKFAQLERFVNGERSLAESQENARRKQQA 642
DB 1650 AKLNUKEQAQIOLIQRAQSIDETTELQEQAKQAKANAPTAKELARRKQQA 1700
RESULT 35
AA034143
ID AA034143 standard; Protein; 2086 AA.
XX
AC AA034143;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #419.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.

```

```

XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT,
PI Yamamoto RT, Xu HH;
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52002.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 5639; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2086 AA;
Query Match 7.1%; Score 228; DB 22; Length 2086;
Best Local Similarity 22.0%; Pred. No. 5.7e-05;
Matches 166; Conservative 106; Mismatches 259; Indels 222; Gaps 35;
QY 25 IINGQIANSNE-----TKES-----TKASEASPSASSVSNFSLSAKNAI 67
DB 1221 ILNKQGSNSDKAADRALQOVSTSDALNGDAKLEAKAAQNLGTLNHTMAORTAL 1280
QY 68 -----SLRDAI--LNKNSPTD-SLSOLEASTS--TSTVTRVAAKDYDEAKSN--F 111
DB 1281 EGOINQATTVYDGVNTYKTNANLTLDGAMNSLQGSINRKDATLRQNTLADDESRRNAYTOA 1340
QY 112 DPAKSGLENAKTLAEYETKADL-----MAALQDMERLANSNHNTEVEVNNIK 161
DB 1341 VTAAGELINKQGTG--NTSKADVDNALNTYTRAKAALNGAENLRNRTKSATNT--INGLP 1396
QY 162 KALEAKDKTI-----DKLNKLYT-----LQONKSLTEVLTQTTDS 196
DB 1397 NLTOLOKMDLKHQVEQAQVAVGNGYKDKGNTLMTAGALRTSIQNDNTTKT-----S 1449
QY 197 ADOIPAINSOLEINKNSADQIILDLERQNTSYEAVLTNGEVIKASSEGKIKGQALQSI 256
DB 1450 QNTLDASDS--NKN-----NYTAVNANAGVNTVNNPNMD--ANAINGM 1490
QY 257 VDAQDQSOAAVILQAOQ-----NNSPDNIA-----ATKELIDAET----- 291

```



```
Db 1491 ANOVTTTAAALNGAQNLAQAATTAKNNLNLTISINNAOKDALTRSIDGATTVAGVNOETA 1550
Qy 292 KVNELKOE-----HGLTDSPLYKKAEEIOISQAOKDIOEIKPSGSDIPYVPSGSA--SA 345
Db 1551 KATELNNAMHSLONGIND-----ETQTKOTQKYLID-----AGPSKKAAYQA 1592
Qy 346 GSAAGAL--KSSN-----NSGISILLDDVDNEMAAILOGFRSIEQFN 388
Db 1593 VMAAKAILTKASGVNDKAAVEOALQNVNSTKTALNGAKLNEAKAAKOTLGTLTHINN 1652
Qy 389 VNNPATAKEL-QAMEAQ-----LTAMSDQLVGADGELPAEIOAIKDALAQAOKPSADGLA 443
Db 1653 AORTALDNEITQATNVEGVNYYKAKAQOLDGAMGLETSIRKDKDTTLOSQNVQDADDAKR 1712
Qy 444 TAMGOVAFPAAKVGGSGAGTAGTYOMNKKLYTAFSSSTSSSYAAALSDGYSAVKTINS 503
Db 1713 TAYSOAVNAATAI-----LNKTAGGNTPRADYERAAQAVTQANTALNG 1755
Qy 504 LYESRSGVQSAISOTANPALIS-----RSVSRSGIESGREGSADASQRAEITIVDS 554
Db 1756 IONLERA-----KOANATFATNMSDLNFKOEALKAQVTSAGRVs-----AANGVEHTA 1804
Qy 555 QTLGDVYSRLQ-----VLDSLMSTIVSNPOANOEIIM-OKLTA-----SISKAPQFGYPA 603
Db 1805 TELNATAMTALMHAIDKAEITKASGVNVNADANKRQAYDEKVTAAENIVSGTPTTILTPSD 1864
Qy 604 VONSADSIQKFAQLEREFDGERSLAESENA 636
Db 1865 VTNATQVTNAKTOL-----NGNHMLEVAKONA 1892
```

RESULT 36

```
AAU37017 standard; Protein; 5795 AA.
ID AAU37017
AC AAU37017;
XX
XX 14-FEB-2002 (first entry)
De Staphylococcus aureus cellular proliferation protein #1187.
XX
XX Antisense: prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
OS
XX
XX MO200170955-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001WO-US09180.
PF
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Haselbeck R, Ohlsen KJ, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
DR N-PSDB; AAS54876.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12610; 511pp; English.
```

```
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from Wipo at
CC ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 5795 AA:
```

Query Match 7.1%; Score 228; DB 22; Length 5795;

Best Local Similarity 22.0%; Pred. No. 0.00022;

Matches 166; Conservative 106; Mismatches 259; Indels 222; Gaps 35;

```
Qy 25 IINGOIASNSE-----TKES-----TKASEASPSASSVSWSLSSAKNALI 67
Db 3498 ILNKOTGNSNRKAAVDRALQOVTSYKDALNGDAKLAEEAKAAKQILNHTINQAORTAL 3557
Qy 68 -----SLRPAI-----LNKSSPTD-SLSQLEASTS--TSYTVRAKAYDEAKSN----F 111
Db 3558 EGOINQATTVGCVNTKTNATLIDGAMNSLOGSINDKDTTLNOMYILDADSKRAAYQA 3617
Qy 112 DTAKSGLENAKTLAEYETKMA DL-----MAALODMERLANSDPSNNHTEEVNNIK 161
Db 3618 VTAAGCILNKOTGC--NTSKADVDAALNTVTRAKAALGAELRRTKTSATNT--INCLP 3673
Qy 162 KALEAKQDPT-----DKLNKLVY-----LONONKSLTEVLKTTDS 196
Db 3674 NLTOLOKDNLKHQVBOAQNVAQVNGVCKDKGNTLNTAMGALRTSIQNDVTTKT-----S 3726
Qy 197 ADQIPAINSOLEINKNSADQIIKDERONISYEVLTNAGVEYKASSBAGIKLGOALOSI 256
Db 3727 QNYLDASDS---NKN-----NNTAVNANGVINTVNPND-ANAINGM 3767
Qy 257 VDAGDOSQAAYLQAO-----NNSPDNIA-----ATKELIDAET----- 291
Db 3768 ANOVTTTAAALNGAQNLAQAATTAKNNLNLTISINNAOKDALTRSIDGATTVAGVNOETA 3827
Qy 292 KVNELKOE-----HGLTDSPLYKKAEEIOISQAOKDIOEIKPSGSDIPYVPSGSA--SA 345
Db 3828 KATELNNAMHSLONGIND-----ETQTKOTQKYLID-----AGPSKKAAYQA 3869
Qy 346 GSAAGAL--KSSN-----NSGISILLDDVDNEMAAILOGFRSIEQFN 388
Db 3870 VMAAKAILTKASGVNDKAAVEOALQNVNSTKTALNGAKLNEAKAAKOTLGTLTHINN 3929
Qy 389 VNNPATAKEL-QAMEAQ-----LTAMSDQLVGADGELPAEIOAIKDALAQAOKPSADGLA 443
Db 3930 AORTALDNEITQATNVEGVNYYKAKAQOLDGAMGLETSIRKDKDTTLOSQNVQDADDAKR 3989
Qy 444 TAMGOVAFPAAKVGGSGAGTAGTYOMNKKLYTAFSSSTSSSYAAALSDGYSAVKTINS 503
Db 3990 TAYSOAVNAATAI-----LNKTAGGNTPRADYERAAQAVTQANTALNG 4032
Qy 504 LYESRSGVQSAISOTANPALIS-----RSVSRSGIESGREGSADASQRAEITIVDS 554
Db 4033 IONLERA-----KOANATFATNMSDLNFKOEALKAQVTSAGRVs-----AANGVEHTA 4081
Qy 555 QTLGDVYSRLQ-----VLDSLMSTIVSNPOANOEIIM-OKLTA-----SISKAPQFGYPA 603
```

DB 4082 TELNTAMTLLKHAIDAKETKASGVNVNADANKRQAYDEKVTAEINIVSGTPTPLTPSPD 4141.
QY 604 VONSADSLQKFAQLEREFEVDGERSLAESOENA 636
DB 4142 VTMAATQVTNAKTQL-----NGNHNLEVAKQNA 4169
RESULT 37
AAG83030
ID AAG83030 standard; Protein: 1095 AA.
XX
AC AAG83030;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:3154.
XX
KM Staphylococcus epidermidis SRI strain; Infection; diagnosis;
XX
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimerly MJ;
XX
DR WPI: 2001-316495/33.
DR N-PSDB; AAH53880.
XX
PT Nucleic acids encoding polypeptides from staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 18; Page 832; 2188bp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55096 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 1095 AA;
Query Match 7.1%; Score 227.5; DB 22; Length 1095;
Best Local Similarity 19.3%; Pred. No. 2.6e-05;
Matches 149; Conservative 136; Mismatches 291; Indels 195; Gaps 31;
QY 16 SSMOSMNPFLINGQISNSETKSTKASESPASSSVSWFSLSAKNALILRLPAIILN 75
DB 181 SKQQAQOETISKQOYVIDNGDATTOQISNAKTIVERALA---LNNAKTGLRADKEELON 237

QY 76 KNSPTDLSQLEASTSTSTVTRVAAKYDEAKSNFDY-----AKSGLEN----- 120
DB 238 -----AVNQLTQNIIDTSGKTSPASIRKYNKASRIQTOIDSKKNKANSILITNDNPQVSQ 290
QY 121 -----AKTAEYETKMAD--LMAALQDMRLANS -DPSNN-HTEEVNNIK 161
DB 291 VTALNKKIAVQPELDKATLMLKNNKENNALVQAKOQLOOVINEVDTQGTDTTDRANNK 350
QY 162 KALEAQKDTIDKLKLVY-----LQONKSLTEVLEKTTDSA-DQIPAINSOLEINKN 212
DB 351 SKKREADELQKQOQIINNNGDATEEQITNETNNVNGAIINAIRKAKNDLADRSQLE---N 407
QY 213 SADQIITKLE-----RONI--STEAVLTNAGEYIKASSPAGIKLGQALQS 255
DB 408 AVNQLIONVDNTGKPKPASIOYQARAQATETQYNNKSEAHQILENSPVSVEVQALQK 467
QY 256 I--VDAGDSQAAVILOAQONNSPDNIAAKRELIDAEFTVNEIKOEHITLQSPLYKKAE 313
DB 468 VEAVQLKVDALHMLQNNKENNS-----ALVTAKNQLQO--AVNDQPLT---- 508
QY 314 EQISQAOKDIOEIKPESGDIPIVGPSSAASAGSAGALKSSNNGRISILK--LDDVDN 370
DB 509 --TGMTQDSINNVVAKRNEAQSAIRNAEAVINNGDATFAQIDSKSKVQALAHINDAKQ 566
QY 371 EMAALALQGRSMIEQFN---VNNP-----ATAKELOAMEAQLTAMSDQLVGADGELP 420
DB 567 QLTADTTE-LQTVAVQOLNRRGDTNNKPRSIINAYNNKAIQSLFQTLTSKANDANAVIQKPI 625
QY 421 AELQAIKDALAQ--LKQPSADGLATAMGOV-----AFAAKVY-----GGSAGTACT 466
DB 626 RTVQEVNNALQOVNQLNQO---LTPAINQLPLSNNDLAKARLLEKNKINQTVQTDGM 681
QY 467 VQNNVQOLKRTAFSTSSSYAALSDGYSA---YKTLNLSYSESSGVSQATIS----- 517
DB 682 TQOSITAYQNAKRVANQENSTALALINNGDADDEQOITTEEDRNOQFTNLQALINQLVYN 741
QY 518 ---QTANPALSRVSRSIGESQGRSADASQRAAETIVBDSQTLGDVYSRLQVLSLMST 573
DB 742 KEPLERAKTALQNNI-----DQVPTDGMTQ-----QSVANYNQKLQIAKKEINMT 786
QY 574 I-----VSNPQANQEIIMOKLTASISKAPQFGYPAVQNSADSLQFAQLERE 621
DB 787 INNVLANNLDVNAIKTNKAEERISNDLTQAKNN-----LQVDTQPLEKIKRQLQDE 838
QY 622 FVDEGERSLAESQF-----NAFRKQAPATQOVUNVANS 653
DB 839 IDQGTITDGMTQDSVQNYNDSLAAIIEKGKYNKLLKRNPTVEQVYESVAN 889
RESULT 38
AAU34320
ID AAU34320 standard; Protein: 2478 AA.
XX
AC AAU34320;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #596.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO2001170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.

```

PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX
XX N-PSDB; AAS52179.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 5816; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2478 AA:
Query Match 6.9%; Score 222; DB 22; Length 2478;
Best Local Similarity 19.6%; Pred. No. 0.00016;
Matches 136; Conservative 114; Mismatches 275; Indels 170; Gaps 26;
QY 14 SVSSNOSNMPIINGQJANSSEETKESRKASPSASSSVSMFSLSAKALLSLDAI 73
DB 535 AVENKMDALQOOVNSQVDSNHYT-----TASIAEY-----NKLQQAADTI 574
QY 74 LNKNSPTDLSQLEASTSTSTVTRVAAKDYDEAKSNDFRKSGLENAKTLAEYETKAD 133
DB 575 LN-----EDANHKTANRASQADIDGLVTKLQAL--IDNQAIAEIDTKAQE 620
QY 134 LMAALQDMERLANSDP-----SNHTEVYNNIKKALEAKQDTIDKLKLVLTQONKS 186
DB 621 KVTAAQOSKKTQDEVALVTKINNDKNNAIAEINKQTAQGVTEKONGIAVLEOD--V 678
QY 187 LTELKTTDSADQIPAI-----NSOLEIKKNSAD-----QITKDLF--QNT 226
DB 679 ITPTVKQAKODITIAVTTTRKQIKSNASLQDEKDVANDKIGETKAKIDAAITNA 738
QY 227 SYEAVLTNAGEVIRKASSEAGIKLQALQSIYDADQSOAVLQAOQNNSPNIATKELI 286
DB 739 QVEAKTKKAIINDIQTPATTAKAALAEFDE-----VYQAOIDQAPLNPPTTNEEV 790
QY 287 DAAETKYNELKQEHGTGLTDSPLVKAEQISOAKQDIOETIKPSGSDIPIVSGSASAG 346
DB 791 AEAERINAAKV--SGV-----KAIEATTTAQ--DLERVK-----821
QY 347 SAAGALKSSNNSGRISILLDDVDNEMAIALQGRSMTEQFN--VNNPATAKEIQAMEQOL 405
DB 822 -----NEBISKIENITSTQTKMDA-----YNEKQAATAKAKQANATVS- 860
QY 406 TAMSQDLVGADGEL-PAEIOAIKDALQALKOPASADGLATPMQGVAFPAAKVGGSGSAGTA 464

```

```

DB 861 NATNEEVAEADAADAAQKQGLHDIOGVKSKQEVADTKSKYLDKINAIQTNQ-----AKVK 915
QY 465 GTYQAMNKQLYKTTAFSSSTSSSYAALSDGYSAKTTINSLXSESRGVO-----513
DB 916 PAADTEVENAMYNTRKQEIQNSN--ASTTEEQAAATEIDTKQOEARTNIDANTNSDVTTA 974
QY 514 -----SAISQIOT-----ANPALSRSVSRGICSGRSADASQRAAEITVRDSOTL 557
DB 975 KDNSTIAINVOQAATTTKSDAKAFIAQKASERKATIEAMNDSTTEEQAA-----KDYVDQ 1030
QY 558 GDVYSRLQVLDSLMASTIVSNPOANQEIIMQKLTASISKAPQFGYPAYQNSADSLQKPAQ 617
DB 1031 AVYTANMIDIDNAAANNVDNAKTTNEATTAIITPDAAVWP-----AAQALADKY-----Q 1081
QY 618 LEREFVUGERSLAEQSNARFRKOPAFIOOVLVNTA 652
DB 1082 AOETFAIDGNGG-SYTEEKAAKAQOQVTEKTTADAA 1115
RESULT 39
AAU37374
ID AAU37374 standard; Protein: 2478 AA.
XX
XX AAU37374;
XX
XX 14-FEB-2002 (first entry)
XX
XX Staphylococcus aureus cellular proliferation protein #1544.
XX
XX Antisense: prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX N-PSDB; AAS5233.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12967; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery

```

CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2478 AA;

Query Match 6.9%; Score 222; DB 22; Length 2478;
 Best Local Similarity 19.6%; Pred. No. 0.00016;

Matches 136; Conservative 114; Mismatches 275; Indels 170; Gaps 26;

```

OY 14 SYSNOSMNPIINGQIASNETKESTKASEASPSASSSSWSFSLSAKNALISRLDAI 73
DB 535 AVEAMKDALQOQVNSGVNSHT-----TASIAEY-----NKLKQOADI 574
OY 74 LKNSSPDLSLSQLEASTSTSTVTVRAAKDYDEAKSNFDTAKSGLENAKTLAEYETKMD 133
DB 575 LN-----BDANHVKTANRASQADIDGLVTKLQAL--IDNQAAIAELDTKAGE 620
OY 134 LMAALQDMERLANSDP-----SNNHTEEVNKKALEAQDQDTIDKLKLVTLQONKS 186
DB 621 KYTAAQOSKKVYQDEVAALVTKINNDKNNAIAEINKQTTAQGVTEKDGIAVLEOD--V 678
OY 187 LNEVTKTDSADQIPAI-----NSOLEINKNSAD-----QIIRKLE--RONI 226
DB 679 IIPYVPAQAKDIIQAVTTRKQOIKSNASLQDEKDVANDKIGKETTKIKIDIAATTTA 738
OY 227 SYEAVLTNAGEVIKASSEAGIRKQALOSIVAGDQSQAVLAQOONNSPDNIATKELI 286
DB 739 QVEAIKTKAIINQINOTTPATTAKAALAEFEDE-----VVQAQIDQAPLMDPTTNEV 790
OY 287 DAEIKVNLKQBHGTLDSPLYKKAEEIQAQKDIOEIKSGSDIPIVFGSGAASAG 346
DB 791 ABAIERINAKV--SGV-----KAIEATTTAQ--DLERVK----- 821
OY 347 SAAGALKSSNNSGRISLLDDVDNEMAAITALQGRSMIBQFN--VNNPATAKELQAMEAQL 405
DB 822 -----NEISKIENTIDSTQTKDA-----YNEVQAATARRAKQANATVS-- 860
OY 406 TMSDQVLGADGEL--PAEIQAIKDALQALQKQPSADGLATMAGQVAFAAKVGGSAGTA 464
DB 861 NATNEVADADAADAAQOGLDIOVYKSKQEVADTKSKVDKINAIQTO-----AKVK 915
OY 465 GTVQMVKQLYTAFTASTSSSSSYAALSDGYSAKYTLNLYSESRSQV----- 513
DB 916 PAADTEVENAVYTRKQEIQNSN--ASTTEKQAAYTELDTKQKQEARNTINDAANTNSDVTTA 974
OY 514 -----SAISQY-----ANPALSRVSRSGETSGRSDASQRAAETIVRDSQTL 557
DB 975 KUNSTAIINQVQAATTKKSDAAAEIQAOKASEKRTAEANNDSTTEQQA-----KDVNQ 1030
OY 558 GUVYSRLQVLDLSIMSTVYSNPQANQEIIMQKLTASISKAPQEGYPAVONSADSLQKFAAQ 617
DB 1031 AVYTAADIDNAAANNVDNAKTNEATTAITPDANVP-----AAKQAIADKV-----Q 1081
OY 618 LREPEYDGERSLAESQENAFKQPAFIOQVLVNIA 652
DB 1082 AQETALDGNNG--STTEKAAAKQOVOTEKTTADAA 1115

```

RESULT 40

ABB71141

ID ABB71141 standard; Protein; 1833 AA.

AC ABB71141;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 40215.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL15244.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Disclosure; SEQ ID NO 40215; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1833 AA;

Query Match 6.9%; Score 219.5; DB 22; Length 1833;

Best Local Similarity 21.6%; Pred. No. 0.00016;

Matches 145; Conservative 105; Mismatches 265; Indels 155; Gaps 26;

```

OY 16 SYSNOSMNPIINGQIASNETKESTKAS--EASPSASSSSWSFSLSAKNALISRLDAI 74
DB 250 SSNSS-----TSIKTSNSTSASNEKVVVTSTSSSTSSSVRRKKEADSVASKEI 299
OY 75 NKNSPDTLSQLEASTSTSTVTVRAAKDYDE--AKSNFDTAKSGLENAKTLAE--YETKM 131
DB 300 KRQYVPAASISH--SNSSTASTASASKQDPTNGOQKALKLELPTTKTAKEAERHK 356
OY 132 ADLMAALQDMERLANSDSNNHT--EEVNNIKKALEAQDQDTID-----KLN-KLVTLQ 181
DB 357 SDIL-----LRLASMDTASNRTAASEALNLOQKLENMEKQDRTYEDKRXKLNLMKEKE 411
OY 182 NQNKSLTEVLKTTSDADQIPAINSOLEINKNSADQIIDQLEQONSYEAVLTNAGEVIKA 241
DB 412 NKGS-----ESELRRKLQAEQICELEMEQNSAKKEIIN----- 446
OY 242 SSEAGIKLQALQSTIVD--AGDQSQAVLAQOONNSPDNIATKEL-----IDAAETRYN 294
DB 447 -----LQAEKDEVDQDFFRDEYKAKTSLOKDEKATKNKRILISFKLKSKDRKE 495
OY 295 ELKQEHGTGLTDSPL--VKKAEQI-----SQAQKDIOEIK--PSGSDIPIVFGSGSA 343
DB 496 TLEQEROSSFNAELSNKIKKLEELRFSNELTRKLQAEAEELRNPGKKAPMLGYLGKST 555
OY 344 SAGSAAGALKSSNNSGRISLLDDVDNEMAAITALQGRSMIBQFN--VNNPATAKELQAMEA 402
DB 556 SA--DAKFTRESLTRG-----SODEPQHLQRELQDSI 586

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 11:22:42 ; Search time 72 Seconds
(without alignments)
1888.765 Million cell updates/sec

Title: US-10-007-693-139

Perfect score: 3204

Sequence: 1 MHHHHHMSPESSVSNQS.....PAFIQVLVNIAISLFGYLS 660

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2943	91.9	647	16	084627 chlamydia t
2	894	27.9	651	16	092787 chlamydia p
3	453	14.1	715	2	046166 chlamydia p
4	272.5	8.5	3381	2	09KX33 streptococ
5	272	8.5	2283	2	08V099 streptococ
6	268	8.4	2271	16	0990Y4 streptococ
7	261	8.1	6713	16	0990Y4 streptococ
8	261	8.1	6713	16	0990Y4 streptococ
9	237.5	7.4	2016	5	09B170 pleistococ
10	235	7.3	2285	9	064046 bacterioph
11	235	7.3	1063	16	031976 bacillus su
12	234	7.3	1063	16	09C886 lactococcus
13	233	7.3	1086	16	09A080 streptococ
14	233	7.3	1344	2	049545 mycoplasma
15	231	7.2	1365	2	049525 mycoplasma
16	230	7.2	2178	2	09KWR3 streptococ

17	228	7.1	1086	2	0938K3 streptococ
18	228	7.1	3072	2	0939N5 streptococ
19	227	7.1	2478	2	091CH2 streptococ
20	226.5	7.1	969	5	09MD19 plasmidum
21	226	7.1	1795	2	091CJ9 streptococ
22	225.5	7.0	1302	2	049547 mycoplasma
23	225	7.0	1404	2	08R3J9 mycoplasma
24	224.5	7.0	1790	3	007380 mycoplasma
25	222	6.9	1327	2	09X7M2 streptococ
26	222	6.9	2478	2	09RL69 streptococ
27	221.5	6.9	924	5	015738 dictyostell
28	220.5	6.9	1627	5	096200 dictyostell
29	219.5	6.9	1833	5	09VM67 dictyostell
30	219.5	6.9	2139	5	007569 entamoeba h
31	219	6.8	1514	5	085Y55 drosophila
32	218.5	6.8	2570	2	09ZFE9 streptococ
33	218	6.8	4776	16	097P71 streptococ
34	217	6.8	1935	5	044934 loliigo peal
35	217	6.8	2297	3	09HGK6 candida alb
36	216.5	6.8	2055	2	085472 abiotrophia
37	215.5	6.7	2124	16	098M03 rhizodinium 1
38	214	6.7	716	2	049526 mycoplasma
39	214	6.7	1072	16	09CF64 lactococcus
40	213	6.6	1489	5	087805 drosophila
41	213	6.6	1794	9	0971A7 bacterioph
42	213	6.6	2186	16	099P80 streptococ
43	212	6.6	962	2	049546 mycoplasma
44	212	6.6	1489	5	09W1R3 drosophila
45	212	6.6	1690	5	044929 drosophila

ALIGNMENTS

RESULT 1
ID 084627 PRELIMINARY: PRT: 647 AA.
AC 084627:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CHLPN 76 kDa homolog.
GN CT622.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX:
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V., Davis R.W.:
RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."
RT Science 282:754-759(1998).
DR EMBL: AE001333; AAC68226.1; -
DR PRC1-2DPAGE: 084627; -
KW Complete proteome.
SQ SEQUENCE 647 AA; 68525 MW; COD1AC2D74473625 CRC64;

Query Match 91.9%; Score 2943; DB 16; Length 647;
Best Local Similarity 93.9%; Pred. No. 7e-110;
Matches 613; Conservative 15; Mismatches 19; Indels 6; Gaps 1;
QY 8 MESSGESVSSNNSNNPIINCOIASNSEETKRSSEASASSVSSWSPFLSSAKNALI 67
DB 1 MESSGESVSSNNSNNPIINCOIASNSEETKRSSEASASSVSSWSPFLSSAKNALI 60
QY 68 SLRDAILNKNSSPTDLSQLEASTSTSTVTVRAADYDEAKSNFDTANSGLLENATLAAY 127
DB 61 SLRDAILNKNSSPTDLSQLEASTSTSTVTVRAADYDEAKSNFDTANSGLLENATLAAY 120

DE 76 kDa protein.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. *Chlamydia*.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94156481; PubMed=7509320;
RA Perez-Melgosa M., Kuo C.-C., Campbell L.:
RT "Isolation and characterization of a gene encoding a Chlamydia
pneumoniae 76-kilodalton protein containing a species-specific
epitope."
RL Infect. Immun. 62:880-886(1994).
DR EMBL; L23921; AAA23117.1;
SQ SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;

Query Match 14.1%; Score 453; DB 2; Length 715;
Best Local Similarity 28.9%; Pred. No. 1.1e-10;
Matches 136; Conservative 94; Mismatches 193; Indels 48; Gaps 14;

OY 11 GPESVSSNOSNMPIINGQIASNSEKSTKA-----SEASP--SASSSVSSMPLSAK 63
DB 262 GGPIDERTERTPADISAOGLSAANKSAEORLAGAKREKSTDSVERKSIIRSAV 321
OY 64 NALISLRD--AILNKNSEPTDSLQLEASTSTVTYRVAADYDEAKSNFTAKSGLENA 121
DB 322 NALMSLADKLGIASSNSSSTSR--ADVDTATATAPPTPEFDYKTKQAOTAYDTIFTS 380
OY 122 KTLAEYETKMDLMAALQMERLANSDPNNHTEVNNIKKALEAKOTID-----K 173
DB 361 TSLADIALVALVLODAVNTIKIDTAID-----EET-----AIAEMETKMDAVKVAQ 429
OY 174 LNKLVTLQNONKSLTEVLKTTSDADQIPAINSOLEINKNSAQIITKDLRONI---SYRA 230
DB 430 ITELKAYASDNQAIIDSLGKLTSPDLQALLQSVANNKKAELKEMDNVVEPKTPA 489
OY 231 VLTNAGEVTKASSEAGIKLQALQIVDAGDOSQAVALQAOONSPNDIAATKELIDAE 290
DB 430 IQASIVDQDATATQTEKDGNAIRDAYFAGQNASGAVENAKSNISINIDSAKAATATAK 549
OY 291 TVNLEKQHTGLTSPLYKKAEBQISOAKDIOELKPS--GSDIPYVGS--GSASAGSA 348
DB 550 TQIAEAKK--FPDSPIQEAQVNTQAEKDKLNKPGSDPPTGTVGSGKQGS 606
OY 349 AGALKSSNSGRISLLDDVDNEMAIALQGFMSIEQFNVNPNPATKELQAMEAOLTM 408
DB 607 IGT-----RVSMLDDAENETASILMSGRQMTIHMTENPDSQAQOELAAQARA 659
OY 409 SDQLVGADGELPAETQAIKDALAQL--KQPSADGLATAMGOVAFPAKAYG 458
DB 660 K---AAGDDSAALADLAQKALEAALGKAGQCGGILNALGQIASAAVVSAG 707

RESULT 4

OYKX33 PRELIMINARY; PRT; 3381 AA.
AC OYKX33;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE StrpA.
OS Streptococcus cristatus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=45634;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CSA;
RA Correia F.F., Allen T.W., Dirienzo J.M.;
RT "High molecular weight serine-rich protein gene (srpA) from
Streptococcus crista";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; U96166; AAF34780.1;
DR InterPro; IPR001899; Gram_pos_anchor.

DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 3381 AA; 321845 MW; C623EE363E74853F CRC64;

Query Match 8.5%; Score 272.5; DB 2; Length 3381;
Best Local Similarity 21.2%; Pred. No. 0.012;
Matches 144; Conservative 130; Mismatches 305; Indels 101; Gaps 19;

OY 9 ESGPESVSSNOSNMPIINGQIASNSEKSTKA-----EASPSASSVSS 55
DB 1378 QSASASANNONSASISVASHASISASOSTSASISASISASOSASVSA 1437
OY 56 WFLSAAKNALSLRDAIILNKNSEPTDSL-----OLEASTSTVTYRVAADYDEAKSN 110
DB 1438 SOSTASVSAOSVSAASANNONNSASISVASHASISASOSTSA-----SISAS 1487
OY 111 PPTAKSGLENAKTLAEYETKMDLMAALQMERLANSDPNNHTEVNNIKKALEAKOT 170
DB 1488 ISASOSASVSAOSASVSAOSOSTSASVSAOSASASANNONNSA--SISASISASQSA 1545
OY 171 IDKLKLVTLQNONKSLTEVLKTTSDADQIPAINSOLEINKNSAQIITKDLRONIS--YE 229
DB 1546 -----SVASOSASASVSAOSOSTSASVSAOSASVSAASANNONNSASISVASHASISASQ 1600
OY 230 AVLTNAGEVTKASSEAGIKLQALQIVDAGDOSQAVALQAO-----ONSPNDIA 280
DB 1601 STSASISASISASOSASVSAOSASVSAOSOSTSASVSAOSASASANNONNSASISVS 1660
OY 281 ATKELIDAEYETVNELKQHTGLTSPLYKKAEBQISOAKDIOELKPSGSDIPYVGS 340
DB 1661 ASOSASISA-----SOSASASISASISASOSASVSAOSASVSAOSASOS-----TSA 1707
OY 341 SASASGAAGALKSSNSGRISLLDDVDNEMAIALQGFMSIEQFNVNPNPATKELQ 400
DB 1708 SVSASOSASASANNONNSASISVA-----SHASIS-----ASOSASASISASISASQSA 1759
OY 401 MEAOLTAMSDQLVGADGELPAETQAIKDALQKPSADGLATAMGOVA-----FAAA 454
DB 1760 VSASOSASAS--VSASOSTSASVSAOSASASASANNONNSASISVSAOSASISASQSA 1817
OY 455 KYGGG--SAGTACTVQMNQVLYKTPAFSTSSSYAALSDGYSAYTTLMSLYSESGVQ 513
DB 1818 SISASISASOSASVSAOSASVSAOSOSTSASVSAOSASASANNONNSA--SISVASQ 1876
OY 514 SA-----ISOTANALRSVRSRGIESGRADSORAETIVDSOTLGVYRLYLDSL 570
DB 1877 SASISASOSTSASVSAOSASVSAOSASVSAOSASVSAOSOSTSASVSA-----SOS 1931
OY 571 MSTIVSNPQ-----ANQEIIMOKLTASISKAPQFGYPAYONSADSLQKPA 615
DB 1932 ASASASNNONSASISVASHASISASOSTSASISASISASOSASVSAOSASVSA--A 1989
OY 616 AOLREFVDCERSLAESQEN 635
DB 1990 SOSTSASVSAOSVSAASN 2009

RESULT 5

OYQV99 PRELIMINARY; PRT; 2283 AA.
AC OYQV99;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Serine-threonine rich antigen.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcaceae; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FR1326;
RA Sharp L.J., Henderson B., Poole S., Nair S.;
RT "Identification of a putative serine-threonine rich antigen from

RT Staphylococcus aureus."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF459093; AAL58470.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 2283 AA; 228866 MW; 9C0991E0E5B24B0 CRC64;

Query Match 8.5%; Score 272; DB 2; Length 2283;
Best Local Similarity 20.4%; Pred. No. 0.0075;
Matches 144; Conservative 142; Mismatches 308; Indels 112; Gaps 20;

QY 10 SCPESSVSSQSSNNPIINQIANSNETKSTKASESPSASSVS---SWSFLSSAKNAL 66
DB 814 TSESSSVSSSTSTSLVNSQSVSSMSDASKSTSLDSISNSSSTPEKSESLSTSDSL 873
QY 67 ---ISRDILINKNSPTDLSQLEA-STSTSTVTVAAKDYEAKNPDPAKSGLENK 122
DB 874 RTSTSLSDSL---SMSTSGSLSKQSLSTSDSASTSOSVSDSTNSISTAEISLESAS 930
QY 123 TLAETKMAADLMAALQDMERLANSDDPSNNHTEEVNNIKKALEAKDTIDKLKLYTLON 182
DB 931 T-----SDSISISNS-----IANSASATSKDSQSTSLSTSDS 966
QY 183 QNKLTEVL---KTDSADQIPAINSOLEINKNSADQIIKIERONISTEAVLTNAGEV 238
DB 967 KSMSTSESLSDSTSTSDSGSLVAGSOSVSTSTSDSM-----STSEI 1010
QY 239 IKASSAGIKIKGALQSIYDAGQSOAAVLQAOQNNSPNIAATKELIDAETKVELNQ 298
DB 1011 VSDSISTSGSL-----SASDSKSMVSSSMSTSGSTSESLSDSOSTSDSDSKSLSLST 1065
QY 299 EHTLTD-----SPLYKKAEEQ-----ISOAKDIOEIKPESGSDIYVPGSGAAGASA 348
DB 1066 SSGSGSTSTSTSSVSTSESGSTSGSMSTSGSDSTSLSTFSDPTSDSKASATASESI 1125
QY 349 AGALAKSNNNGRI--SLLDVDNEMAAIALQGRSM-----IQGFVNNPATKKEIQ 399
DB 1126 SOSV-STSTSGSVSTSTSLSTSNSESTSTSMSPDSTSLSTSESDSTSDSTSEIS 1184
QY 400 AMAQOLTASMDQVAGDELPAETQAIKDALQALQKOPASDGLA----- 443
DB 1185 GSESTISISENSSTGDSSEKASAFLESLESTSESTSESLSGSTSDSTSLSDSNES 1244
QY 444 -----TAMQVAPAAKVGSGAGTAGTYQNNVKQLYKTAFTSSSSYAALSDG 494
DB 1245 GSTSTSLMSTSGSTSTSTSGASTSTYKSESVSTSLSTSTSLSDSTSLSTSLSDS 1304
QY 495 YSAVKTLLNLSIESKSGVSAISQITANPALSRVSRSGIESQGRSDASORAAETTVRDS 554
DB 1305 TSGSKS-NSL-SASMS-TSDSISTRKSESLASTSLSGSTSESESGSTSSASOS--DS 1358
QY 555 QRTGADVSRLOYDLMSRTIVSNPOANOEIMQKLFASISKAPQFPPYAVONSDSLQKR 614
DB 1359 TSMSS--LSMSQSTSGSTSTSTSLSDSTSLSLASMSQGVDSNSASQASSTSTSLIS 1416
QY 615 AAQLEREYVDEGERSLAESQENAFRKQAPFQVLYVNIASFGLYS 660
DB 1417 TSSSDQSTSTYTSQSTSGSESTSTSTSLSDSTSIKSTSGSGSTS 1462

RESULT 6
Q990Y4 ID 0990Y4 PRELIMINARY; PRT: 2271 AA.
AC 0990Y4
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical protein SAV2654.
GN SAV2654 OR SA2447.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Staphylococcus aureus (strain N315).
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21111952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Ikeno J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hirumatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003366; BAB58816.1; -
DR EMBL: AP003138; BAB43752.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2271 AA; 227844 MW; 7C2A7040D6C8289D CRC64;

Query Match 8.4%; Score 268; DB 16; Length 2271;
Best Local Similarity 20.6%; Pred. No. 0.011;
Matches 146; Conservative 132; Mismatches 271; Indels 160; Gaps 22;

QY 10 SCPESSVSSQSSNNPIINQIANSNETKSTKASESPSASSVSWSFLSSAKNALISL 69
DB 1134 SGGVSTSTSLSTSNSESTSTSVSDSTSLSTSESDSTSESTSTSDSLSEALISSESTSI- 1192
QY 70 RDAILINKNSPTDLSQ-----LEASTSTSTVTVAAKDYDAKNFPAKSGLEN 120
DB 1193 ---LSESNSTDSSEQSAFAFLSESLSESTSESVSSSTSESLSDSTSEEG-ST 1247
QY 121 AKTLAETKMAADLMAALQ-----DMERLANSDDPSNNHTEEVNNIKKALEAKDTID 172
DB 1248 STSLNSTSGASISSTSTSLSESTSTFKSESVSTSLSMSTSTSLNSTSLSTSLSDSD 1307
QY 173 KLNKVLTLONONKSLTEVTKTDS-----ADQIPAINSOLEINKNSADQIIKPL 221
DB 1308 -----SKSDSLSTSMSTSDSTISKSDSLSTSTSLSGSTSEESDSTSSSEKSDS 1358
QY 222 ERONISYE-----AVLTNAGEVTKASSEAGIKIGAL-QSIYDAGQSOAAVLQAOQNN 275
DB 1359 TSMISMSTSGSTSTSTSTSLSDSTSLSLASAMNQGVDNSAOSASSTSTSTNS 1418
QY 276 PDNIAATKELIDAETKVELKOEHTGLDPSLYKKAEEQI-----SOAKDIQ 324
DB 1419 ESDSQST-STYTSQSTSGSESTSTSTSLSDSTSIKSTSGSGSTSTASISLSESDSQ 1477
QY 325 EIKPESGSDIPIVPGSASAGSAGALKSSNNNGRI--SLLDVDNEMAAIALQGRSMI 384
DB 1478 SITSASE-----STSEASTSLSDSTSTSMSSAS- 1508
QY 385 EGFVNNPATKELIQAMEAQLTAMSDQVAGDELPAETQAIKDALQALQKOPASDGLAT 444
DB 1509 TSTSLSNMSASAE---SDSSSTSLSDSTASAMQSSSDSQSTASLSLSDSLSTSMRMET 1565
QY 445 AMQVAPAAKVGSGAGTAGTYQNNVKQLYKTAFTSSSSYAALSDGYSAYKTLNLSL 504
DB 1566 ---IASLSTSVSTSESGS-----TSESTSESDSTSTSLSDSQSTSR----- 1603
QY 505 YSESRSGVSAISQITANPALSRVSRSGIESQGRSDASO-----RAAETVYDSDQLD 559
DB 1604 -STASAGSASTSTSD---SRSTASSTSMRTSTSDSQMSLSTSTSMKSDTSLSD 1659
QY 560 VVS-----RLQVLDL-----LMSTIVSNPOANOEIMQKLTAST 593
DB 1660 SVSDSTSDSTSGSMVSTSLSDSTSTSTASVMAASISDSQSMESVNDSSVSR 1719
QY 594 SKA-----PQFGYPVAVONS-----ADSLQFAAQLEREYVDEGERSLAES 632

Db 1720 SNSESDSKMSGSTSVSDGSLVSTLRKSESVSSSLSGSQMSDS 1768

RESULT 7

099054 PRELIMINARY; PRT; 6713 AA.
AC 099054;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2002 (TREMBLrel. 17, Last sequence update)
DE Ebha protein.
GN EBHA OR SA1267.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OC NCBI_TaxId=158879;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.T., Nagai Y., Lian J., Ito T., Kanamori M.,
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hiraekawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shibata T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003133; BAB42527.1; -.
KW Complete proteome.
SQ SEQUENCE 6713 AA; 722339 MW; AF6FEDE226BE488 CRC64;

Query Match 8.1%; Score 261; DB 16; Length 6713;

Best Local Similarity 20.5%; Pred. No. 0.078;
Matches 150; Conservative 133; Mismatches 237; Indels 210; Gaps 31;

QY 10 SGPESVSSNQSMMNPINQIASNSETKSTKASESPSASSVSSWSFLSSAKNALISL 69
DB 5610 NCDQNLANKKDKANAFVNSLNLGNLQOQO-----DLAKHAINNADVSVTDIVNNQIDL 5663
QY 70 RDAIILNKSSPTDSLSQL---EASTSTSTVTVRAAKDYDEAKSNEDTAKSGLENATLAE 126
DB 5664 NDAM-----ETLKLVLNDNIPNAEQVYVNYQND--DNAKTNFDDAK----- 5702
QY 127 YETKMAIDLMAALQDMERLANSPPSNHNEEVNNIKKALEAQKDTIDKLN-----KLV 178
DB 5703 -----RLANTLLNSDNT--NVNDINGAIOAVNDAIHNLNGDQRLQDAKDK 5745
QY 179 TLQONKSLTEVLKTTDSA-----DOIIPAINSOLE-----INKNSADQIIKDLERO- 224
DB 5746 AIOSTINQALANKLKEIEASNATDQDKLAKNKAEEELANSIINNINKATSNQAVSQVQTAG 5805
QY 225 NISYFAVLNAGEVIRKASSEAGIKIGQALQSIYVDAGDOSQAVALQAOQNSPDNTAAATKE 284
DB 5806 NHAIEQV--HANEIRKAKIDAKDKVDQVQALIDEIDR-----NPNLTIDK 5848
QY 285 LIDAEETVNE--LKEHRTGLTDS---PLVKAEEQISQAOKDIOEIKPSGSDPIYVPSG 340
DB 5849 EKQALKDRINQILOQHNDINNALKKEIEQAQALQALQIDQIKDLVKAKE----- 5900
QY 341 SNAASGSAAGALKSSNSGRISLLDDVDVN--EMAAIALQGRSMIEQF-----NVNNP 392
DB 5901 -----AKODVVKQVQALIDEIDQNPDLTDEKQALKDRINQILOQHNGINNAA 5948
QY 393 ATAKEIQAIEAQLT-----AMSDOLVYAGDGLPAETQATK 427
DB 5949 MKKEIEQAQKQALQALKEIKLVAKAKENAKODVDKQVQALIDEIDQNPDLTDEKQALK 6008
QY 428 DALAALQKPSADGLATAGQVAFAAKVGSGSAGTACTVQNVKQLYKTAFTSSSSSY 487
DB 6009 DRINQILOQHND--INNATKKEIEQAQ-----AQALQALQ--DIKDLV----- 6050

QY 488 AALSDGYATYKTLNSYSESRSGVQSAISQTPANPALNSVSRSGIESGRSGADASORAA 547

DB 6051 --AKEDAKNAIK--ALANAKKQDQINS-----NPDLPPEQAKAL-----KEIDEAKKA 6095

QY 548 ETIVRDSQTLGDVYSRLQV-LDSLMSITV-----SNPQANO-----EETM----- 586

DB 6096 LQNVENAQRIIDQNLNGLNGLDIDIRNTIHWEDQPAVNEIFEATPEQILLVNGELIVHRD 6155

QY 587 -----OKLTASISKAPQFC-----YPAVONSADSLQKFAAQLEREFVDGERSLAESQENA 636

DB 6156 DIITEQDILAHINLIDQLSAEYIDTPSTATISDSL---TAKVEVILLDSKVIVNVPKV 6212

QY 637 FRKQAPAFIQO 646

DB 6213 VERELSVYKQ 6222

RESULT 8

093186 PRELIMINARY; PRT; 6713 AA.
AC 093186;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein ebha.
GN EBHA OR SAV1434.
OS Staphylococcus aureus (strain Mu50 / ATCC 700659).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OC NCBI_TaxId=158878;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hiraekawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shibata T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003362; BAB57596.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 6713 AA; 722306 MW; BBCC0536AC341BF5 CRC64;

Query Match 8.1%; Score 261; DB 16; Length 6713;

Best Local Similarity 20.5%; Pred. No. 0.078;
Matches 150; Conservative 133; Mismatches 237; Indels 210; Gaps 31;

QY 10 SGPESVSSNQSMMNPINQIASNSETKSTKASESPSASSVSSWSFLSSAKNALISL 69
DB 5610 NCDQNLANKKDKANAFVNSLNLGNLQOQO-----DLAKHAINNADVSVTDIVNNQIDL 5663
QY 70 RDAIILNKSSPTDSLSQL---EASTSTSTVTVRAAKDYDEAKSNEDTAKSGLENATLAE 126
DB 5664 NDAM-----ETLKLVLNDNIPNAEQVYVNYQND--DNAKTNFDDAK----- 5702
QY 127 YETKMAIDLMAALQDMERLANSPPSNHNEEVNNIKKALEAQKDTIDKLN-----KLV 178
DB 5703 -----RLANTLLNSDNT--NVNDINGAIOAVNDAIHNLNGDQRLQDAKDK 5745
QY 179 TLQONKSLTEVLKTTDSA-----DOIIPAINSOLE-----INKNSADQIIKDLERO- 224
DB 5746 AIOSTINQALANKLKEIEASNATDQDKLAKNKAEEELANSIINNINKATSNQAVSQVQTAG 5805
QY 225 NISYFAVLNAGEVIRKASSEAGIKIGQALQSIYVDAGDOSQAVALQAOQNSPDNTAAATKE 284
DB 5806 NHAIEQV--HANEIRKAKIDAKDKVDQVQALIDEIDR-----NPNLTIDK 5848
QY 285 LIDAEETVNE--LKEHRTGLTDS---PLVKAEEQISQAOKDIOEIKPSGSDPIYVPSG 340

```

Db 5849 EKOALKDRINLIQOQHNDINNALTKEEIEQAQALQALQDIDKDKAKED----- 5900
QY 341 SAASAGSAGALKSSNNSCRISLLDDVDN--EMALIALOGFRSMIEOF-----MVNNP 392
Db 5901 -----AKQYDVKQVQALIDEIDONPNULTKEKQALKDRINLIQOQHNGINNA 5948
QY 393 AFAKELQAMEAOLT-----AMSDQLVGADGELPAEIOAIR 427
Db 5949 MKTEEIEQAQALQALKEIKDLVKARENAKQDVQKQVQALIDEIDONPNULTKEKQALK 6008
QY 428 DALAQALQPSADGLATAMGQVAFAAKYGGGAGTAGIVQMNKOLYKTAFESTSSSY 487
Db 6009 DRINLIQOQHND-INNAAMTKEEIEQAK-----AQALQALQ-DIKDLVK----- 6050
QY 488 AALSDGYSAKYTLNLSYSESRGVQSAISQATNPALSRVSRSRGIESQSDASORAA 547
Db 6051 --AKEBAKKAIRK--ALAKAKRQDINS-----NPDLTEPQAKAL-----KEIDEAKRA 6095
QY 548 ETTVRDSQTLGDVYSRLQV-LDSLMSTIV---SNPQANO-----EEM----- 586
Db 6096 LQNVENAOFTIDQINRLNGLDIDIRNTHWEVDEQPAVNEIFEATPEQILVNGELIYHRD 6155
QY 587 -----OKLASTASKAPOFG-----YPAVQNSADSLQKFAQLEREVDGERSLAESQENA 636
Db 6156 DITTEDDILAHNLIDQLSAEVIDTPSTATISDSL--TAKVEVTLLDGSKVIVANPVKY 6212
QY 637 FRKOPAFIOQ 646
Db 6213 VEKELSVKQ 6222

```

RESULT 9
Q9BITO PRELIMINARY: PRT: 2016 AA.

```

AC 09BITO:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Fibroin 3 (Fragment).
OS Plectreurys tritaxis (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Haplogynae; Plectreuridae; Plectreurys.
OX NCBI_TaxID=33319;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-21179804; PubMed-11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
RT Fibroin Sequences.";
RL Science 291:2603-2605 (2001).
DR EMBL; AF350283; AAK30612.1; -.
FT NON_TER 1
SQ SEQUENCE 2016 AA; 196162 MW; F53AA7DB189826F1 CRC64;

```

Query Match 7.4%; Score 237.5; DB 5; Length 2016;
Best local Similarity 22.2%; Pred. No. 0.15;
Matches 157; Conservative 110; Mismatches 316; Indels 125; Gaps 24;

```

QY 14 SVSSNGSSNNPITNGQIASNEFKESTKASEASPS-----ASSSV 53
Db 648 STSIAKOTAASTASTAATSTTQTAATTSASTAASSQTVOKASTSSASTAAQOTGSSSV 707
QY 54 SSWFSFSSAKNALISLDAILINKNSPPTDLSQLEASTST--STVIRVAKDYDEKSNF 111
Db 708 QNGQ-SSSASSSVSDI-----SDSLTSLQSEFTSAFGTSEAEKQSTAEVAQS 762
QY 112 DPAKSGLENAKTLAEYETKADLMALQDMERLANSPPSNHTEEVNNIKKALEAQDIT 171
Db 763 TVAQGLI-----DYSQSSALATAVANAVSQKQSSSAVA-----RAIAYAITTY 808
QY 172 DLINKLYT--LQONQKSLTEVL-----KTDSADQIPAINSOLEINKNSADQITKDE 222

```

```

Db 809 LKTRITITITTTQTVKSPASAISSSLSTARPAISSANAYEQTQSSAAASAAQ--SSEYQ 867
QY 223 RQNISEAVLTVNAGVEIKASSEAGIKQLQALQIVDAGQOSQAAVYQAOQNNSPDNIAT 282
Db 868 TQNTQSSA-----SASSDASTSY-QTQSYSDASAAVA-----ESTSANOQS 912
QY 283 KELIDAEFKVNEKQEHGTGLDPSLVKKAEEQISOAQDI----- 323
Db 913 TQSSAAASSSTNGAYQSOOSYIDASTVSSASANTAOFTYQVITPDNTYFAESTLSTLIQ 972
QY 324 -QEIKPSGDIPIVGPSSGASAGSAGALKSSNNNG-----RISLLDDVDNEMAAIA 376
Db 973 EOPNSKFGSTPLVTAREYASAMARATALLIGFDSGTGTALESAAVAVAVSNVDYASATY 1032
QY 377 LOGFRSMIEQFNVNP--ATAKE---LQAMEAQLTAMSDQLVGADGELPAEIOAIRKDAL 430
Db 1033 ARALIAFASIVLNINQIFPASPASEALYLPAPAMIASLHAFKGSF--SSSAFALANSISPST 1091
QY 431 AQALQOPASADGLATAMGQVAFAAKYGGGAGTAGIVQMNKOLYK-----TAFSSTS 483
Db 1092 AITSAOSSSVSAGASSGQSYDTSSVSSASSEATESSSVFDTYQATQIESAAMAAAS 1151
QY 484 SSSYAAALSDGYSAKYTLNLSYSESRGVQSAISQATNPALSRVSRSRGIESQ--GRSAD 541
Db 1152 SSATVDQFSESSASSAASAFSEOTSYDISDLSASATTAASASSAYESQFSDASSG 1211
QY 542 ASQRAETTVRDSQTLGDVYSRLQVLDLSMSTIVSNPQANOERIMOKLTAS----ISKAP 597
Db 1212 SSAAASASSQNSYDIDALYSA-----SSAAASAAASAVAELEFSDASSSSAAVAVASQ 1266
QY 598 QFGYPAVQ--NSADSLQKRAQ--LEREPVDGERS-----LAESQENAF 637
Db 1267 QGSYDTSDFSSASASAAAASAYESKFLDASSSSSAAAASQOQSSY 1314

```

RESULT 10
ID 064046 PRELIMINARY: PRT: 2285 AA.

```

AC 064046:
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative transglycosylase.
DE YOMT.
GN YOMT.
OS Bacteriophage SPBc2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=66797;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Manuel C.,
RA Karazmata D.;
RT "The complete nucleotide sequence of the Bacillus subtilis spBc2
RT prophage.";
RL Submitted (AUG-1997) to the EMBL/Genbank/DBD databases.
DR EMBL; AF020713; AAC13005.1; -.
DR HSSP; P00718; 153L.
DR InterPro; IPR002886; Peptidase_M37.
DR InterPro; IPR00189; SLT domain.
DR pfam; PF01451; Peptidase_M37; 1.
DR pfam; PF01464; SLT; 1.
DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
SQ SEQUENCE 2285 AA; 252310 MW; FF602C227754B357 CRC64;

```

Query Match 7.3%; Score 235; DB 9; Length 2285;
Best local Similarity 19.3%; Pred. No. 0.22;
Matches 159; Conservative 141; Mismatches 312; Indels 210; Gaps 33;

```

QY 10 GGPESVSSNQSSNNPIT--NGQIASNE-----TREKTSASEASPSASSSVSSWFLSS 61
Db 56 SAIDTYQKMLKSYNGTVKSTYIKNADGSVEKLTQGYKKNGEILLRETKRIINN----- 109
QY 62 AKNALISLDAIILINKNSPPTDLSQLEASTSTSTSTVIRVAKDYDEKSNFD-----TAK 115

```

```

Db 110 -RNTALKEOETEVNKLTOATEKLGQVOKKTVQVRNLQGOPTKVVQKNRHHGFDIYVTTDPK 168
Oy 116 SLENAKTLAEYETKMAIDLMAALQDMERL-----ANSDESNHTEEVN 158
Db 169 TNSTSSKTTTNTYDQORRAIEOLKQLEKLRQGIYTDTTISSLGKKINTAQSAQOIEALQ 228
Oy 159 NIKKALEAKQDITDKLNL-----VTLQONKSLTEVLKTTDS---ADQIPAI 203
Db 229 NRIRMLDOKSAAVAKNNELKTIELYQROAQVNVQNLMTFRYSSGSSNRQAVQDYLVNAV 288
Oy 204 NSOLEFNKNSAQDIKKLERONISYEAVLTNAGEVITKASSEAGITLQALQSI---VDAG 260
Db 289 NS-LNVSTGS--NNIRSOQLSLNMQPRELANSNOTAANOASSFGAELTQTFKSMSTYLLSG 346
Oy 261 D-----OSQAAVLQAQNNSPDNIAATKELIDAAETKVNELKOEHTGLTDSPL 308
Db 347 SLFYCAISGLKEMVSALEIDTLMTN-----IRVMNEPKYKNELQESIDLDGDT-L 398
Oy 309 VKKA-----EEOISQAOKDIOEIKPSSDPIYVPPSSASAGSAGAL 352
Db 399 SNKITDILQMTGDFGRMGFDESELSLTUKTAQVLO---NVSDDLPPDVTNLTAAMLNF 454
Oy 333 K-SSNNSGRISILLDDVDNEMAAILQ-----GFRSMI--- 384
Db 455 NIAANDSISIADKLNEVDNNVAVTTLDLANSIRKAGSTAFGEVLNDLIGYTTAIASTP 514
Oy 385 -EQFV-----NNPATKELQAMPAQLTAMSDQVAGDELPAITQAIKDAL 430
Db 515 RESGIVGNSIKTIFARIGNNOSSIKALEQIGISVKTAGEKKA-SDLISEVAKKMDTL 573
Oy 421 AQAALQKPSADGLA-----TAM---GOVAFPAKVGGSAGT-----AGTVOMV 471
Db 574 SDAQQNTSICVAGYIQLSFRFAMNNNSIAQNAKTAANTSGSAMSEQKAKDADLQARY 633
Oy 472 KOLK--TAFSTSSSYAALSDGSAY-KTNSLYSESRGVS-----AISQTA 520
Db 624 NKLQNNTEFAIADDAF---ISDGLIEFTQAAGSLNASTGVIKSGFLLPPLAAVSTP 690
Oy 521 NPALS-----SVSRGIESGGRSDASQARAETIVRD---SQTIG 558
Db 691 TLLSKNTRTLASSLILCTFRAGOEFTLATAGLEAGMTRAASRVLKALRKALVSTLVG 750
Oy 559 DVYSRL-OVLDSLMTIYSNPOANOELIMOKLTASISKAPQPGYPAVONASDLOKFAAQ 617
Db 751 GAFALGALAESLISFPAKAKAD-----FBSQOQNVNVAITTNKDSITKLIQ 801
Oy 618 L-EREVDGERSLAESQENAFKQPAFTQOVLVNIASLFGY 658
Db 802 YKELQVKESRSITSDQEYLIQ---VTQOLAQTFPALVKGY 840

```

```

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Hentaut A.,
RA Hilbert H., Holsappel S., Hosono S., Huijlo M.F., Itaya M., Jones L.,
RA Joris B., Karameta D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kunita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noack M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle D., Porrolik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
RA Viari A., Wandut R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RA "The complete genome sequence of the gram-positive bacterium Bacillus
RA subtilis."
RA Nature 390:249-256(1997).
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RA Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: 299115; CAB14053.1; -.
DR HSPSP; P00718; 153L.
DR InterPro: IPR002886; Peptidase_M37.
DR InterPro: IPR00189; SLT_domain.
DR Pfam: PF01451; Peptidase_M37; 1.
DR Pfam: PF01464; SLT; 1.
DR PROSITE: PS00922; TRANSGLYCOSYLASE; 1.
KW Complete proteome.
SQ SEQUENCE 2285 AA; 252310 MW; PF602C227754B357 CRC64;

Query Match 7.3%; Score 235; DB 16; Length 2285;
Best Local Similarity 19.3%; Pred. No. 0.22;
Matches 159; Conservative 141; Mismatches 312; Indels 210; Gaps 33;

```

QY 353 K-SSNNGRISLLDDVDNEMAAIALQ-----GFRSMI----- 384
Db 455 NIAANDSISIDAKLNEVNNYAVTTLLDLANSIRKAGSTAFGEVLENDLIGYTTAIASTT 514
QY 385 -EOPNV-----NNPATAKELOAMEAOLATMSDOLVGADGELPAEIOAIKDAL 430
Db 515 RESGIVGNSLKTIRIARIGNNOSSIKALEFOIGISVKTAGEAKSA-SDLISEVACKWTL 573
QY 431 AQAALQPSADGLA-----TAM--GOVAFPAKVGGSAGT-----AGTVOMNV 471
Db 574 SDAQONTSIGVAGIYQISRFNAMNNFSLAQNAKTAANSIGSAMSEQOKYADSLQARV 633
QY 472 KOLYK--TAFSSTSSSTAALSDGYSAY-KTLNSLYSSRGVGS-----AISQTA 520
Db 634 NKLQNNFTEFAIYASDAF---ISDGLIEPTQAAGSLNASTGVISVCFPLPLLAAVSTA 690
QY 521 NPALSR-----SVSRSGIESOGRSADASORAETIVR---SQTG 558
Db 691 TLLSKNRTTLASSLILGRAMGOETLATAGLEAGKTRAAVASRYLKTLRGLVSTLYG 750
QY 559 DYSYRL-QVLDLSMSTIVSNPOANEIMQKLTASISKAPQGPAYONSADSLQKFAAQ 617
Db 751 GAFALGMALESLSISFAEAKKAKDD-----FEQSQQTNVEAIFTNKDSTDKLIQ 801
QY 618 L-EREFVDEGRSLASQENAFKQPAFTIOQVYNITASLFSGY 658
Db 802 YKELQKVESRSLTSDDEOYLQ---VTQQLAQTPPALVKGY 840

RESULT 12

Q9CH86 PRELIMINARY: PRT: 1063 AA.

AC Q9CH86: 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Unknown protein.
GN Y1HD OR L10852.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacilli/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Boletín A., Wincker P., Manger S., Sorokin A., Malarme K.,
RT Weissenbach J., Ehrlich S.D., Sorokin A.,
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006319; AAK04950.1; -.
DR InterPro: IPR001899; Gram pos. anchor.
DR PRINTS: PR01608; BACINVASINC.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 1063 AA; 104547 MW; 5727ADA9C69669A1 CRC64;

Query Match 7.3%; Score 234; DB 16; Length 1063;

Best Local Similarity 22.4%; Pred. No. 0.094; Indels 160; Gaps 29;

Matches 160; Conservative 108; Mismatches 285;

QY 13 ESSVSNQSNMPT-----INQIASNSETKESTKASEAS--PSASSVYSWSFL-- 59

Db 146 DSSSNDSNNSNISLSSNADVDASVQSSTSSGVSSESAIDSGIASVSQSEML 205

QY 60 -----SSAKNAISLRLDALNKNSSPTDSLQ-----LEASTSTVTVAAKD-YDEAK 108

Db 206 YCNSSASASAAVASFATLATNPMSVPMLTQALAAAPATISGAILNTTTLCDLVNCAI 265

QY 109 SNFPAKSGLENA-KTLAYETK-MADLMAALQDMERLANSNPNHTEVNNIKKALA 166

Db 266 ST--VGISGLANIFSTLGTENIPGMTTAAALNGEIVNI-----VGNIOEA--- 311

QY 167 OKDTIDKLNKLTATLONOKMSLFEVLTQDSDAQIPAINSOETINKNSADQIIKLERONI 226
Db 312 -----AANPGAFILNEIKS-----AGLDVSOQIPLVGOIAAENAI-----PSM 350
QY 227 SYEAVLT--MGEVITKASSEAGIKLQALQSIYVDQDQSOAAVLOAQONNSPDNI--AAT 282
Db 351 SPAAMLFLENPFTTIPGLSIPGASL--VLSPLVASITVTSQIVN-QLNTTISNALGVN 407
QY 283 KELIDAEFTKYNELKQETHGLTDSPLVKRAEBOISOAQKDIOEIKPSSGDIPIVPGSG-- 340
Db 408 FDLDTLVLSQGDVINYLAGLVNNSAINRV-QQIMSQ-----LSPTISNIPLVGTVNN 461
QY 341 -----SAASGASGALKSSNNSGRISLLDDVDNEMAAIALQGFRTIEQFNV-- 389
Db 462 VLSPTLNNLTGASLGEVA-----NLGVSSLDQVNNISLGNLISLSTALATIENTLQ 514
QY 390 -----NNPATAKE-----LOAMEAOLATMSDOLVGADGELPAEIOAIKDALQALQK- 436
Db 515 NSLNSFGNLPAGASDILNOVLQONALNINIVESATGIYNNLPQ-LGAIENGLSTISQI 573
QY 437 -----PSADGLATAMGOVA-----PAAKVGGSAGTAGTVOM 469
Db 574 PNINNFVNNALNGITTTIINSLTSPVGASTVPNNSANSOSSSSSSAASSSTSS 633
QY 470 NKKLYKT-----AFSSTSSSTAALSDGYSAYKTLNSLYSESRGYSQSAISQTANPA 523
Db 634 NVSNVTSSNSSEANTSSSTSNASSSSSSEGS6SA-----SSSSSESSVASSSSVDSOS 688
QY 524 LRSVSRSGIESOGRSADASORAETIVRDSQTLDDVYSRLQVLNLSMSTIVSNPOANE 583
Db 689 SSAGVNSSSSSAEGSSASSNSSESVASSS-----VSSOS---SSAGVMS 734
QY 584 EIMQKLTASISKAPQGPAYONSADSLQKFAOLEREFDERSLASEQENA 636
Db 735 SSSSSSSASSNSSEGSVASSSVDSOSSAGVNSSSSSAEGSSASSNSNS 787

RESULT 13

Q9A0N0 PRELIMINARY: PRT: 1086 AA.

AC Q9A0N0: 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical protein SPY0697.
GN SPY0697.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacilli/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Plameaux C., Sezate S., Suvoirov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: AE006522; AAK33654.1; -.
KM Hypothetical protein: Complete proteome.
SQ SEQUENCE 1086 AA; 115601 MW; 87592EB52CDD5ADB CRC64;

Query Match 7.3%; Score 233; DB 16; Length 1086;

Best Local Similarity 21.4%; Pred. No. 0.11; Indels 212; Gaps 34;

Matches 168; Conservative 116; Mismatches 288;

QY 14 SVSNQSNMPTIINQIASNSETKES-----TKASEA-----SPSASSSVS 54

Db 355 SMSIGATITIKYTAQVAVSTMTKADMVARLSQGLYKAKSYIYGVNGAISLSTAAITIA 414

RC STRAIN-MH56;
 RX MEDLINE-95369882; PubMed-7543881;
 RA Jensen L.T., Ladefoged S., Birkelund S., Christiansen G.;
 RT "Selection of Mycoplasma hominis PG21 deletion mutants by cultivation
 in the presence of monoclonal antibody 552.";
 RL Infect. Immun. 63:3336-3347(1995).
 DR EMBL: U21962; AAA81013.1;
 SQ SEQUENCE 1365 AA; 154983 MW; 847A04992410867F CRC64;

Query Match 7.2%; Score 231; DB 2; Length 1365;
 Best Local Similarity 21.5%; Pred. No. 0.17;
 Matches 149; Conservative 137; Mismatches 288; Indels 120; Gaps 28;

QY 14 SVSSNOSNMPIINGQIASNETKESTKASESPSSASSVSSKPLSSAKN-----ALIS 68
 DB 246 SMQSAKSSIDAKV-AEITFKLETFNKDEKAFNELKQTRNQIOEFINTKNNPNYSELIS 304
 QY 69 LRDAIINKSSPTDSLQLEASTSTSTVTRVAK-DYDE-----AKSNPTAKSULENA 121
 DB 305 QLTSKKDSKNSVTDSSNKSIDIESANTELKQALAKANADKQADNLAKSIEQLNNSVSN 364
 QY 122 KTL-AEYETKMAIDMALADMER--LANSDPNNHTEEVNNIKALEAQKOTIDKLNKL 177
 DB 365 NTLASALTIDKDNITQAKTELEKEVQKADQAIKSNNNTASMQSAKSSLDKVAETIK--KL 422
 QY 178 VTLQONKSLTEVLTQTTSDAQIIPAINSOLEINKNSAD--QIIKLERONISYEAVL--T 233
 DB 423 EEFNKDEKAFNELKQTRN-----QIOEFINTKNNPNYSELISQLTSKRDSKNSVTDSS 477
 QY 234 NAGEVIRKASSEAGIKIGQALQSTIVDAGDOSQAVALQAOONSPDN--AATKELIDAEIK 292
 DB 478 NMSDISANTELKQALAKANADKQADNLAKS--IKEQLNNSVSNNTLSAKLTD-----K 531
 QY 293 VNELKQHTGLTDSPLVKKAEEQISOAKDIOEIKPSGSDIPVGPSSAASAGSAGAL 352
 DB 532 DNTIQAKTEL-----EKEVQKADQAIK-----SNNTASMQSAKSSL 568
 QY 353 KSSNNSGRISLIDVDNEMAAIATAGFRSMTEQF--NVNPNATKELQAMEQLTAMS 409
 DB 569 DAKVAETIKLETFNKDEKAFNELKQTRNQIOEFINTKNNPN--NYSSELISQLTSKR 624
 QY 410 DQLVG-ADGELPAEIOAIDALQALQKPSADGLATAMQOVAFAPAAKVGGSAGTGYO 468
 DB 625 DSKNSVTDSSNKSIDIESANTELKQALAKANADKQVQ--DNLAKSIEQLNNSVSNNTLS 682
 QY 469 MNV-----KQLYK-----TAFSTSSSYAALSDQSYAKTILN-- 502
 DB 683 AKITDKNITQAKTELEKEIQKQANQAIKSNNNTASMQSAKSSLDKVAETIKLETFNKD 742
 QY 503 -----SLYSESRSGVSAISQTA-NPALSRVSRSIGIESQGRSADASQRAAETIVDSQT 556
 DB 743 KEAKFNEIKQTRNQIOEFINTKNNPNYSELISQLE-----TSKRDSKNSVTDSSN 792
 QY 557 LGDVRLOYLDLSMTIYSNPQANOEIMOKITASTISAPQPGYAVQNSADSLQKFAA 616
 DB 793 KSDIESANTELQALNTAKAKSSIDNELRPLKNDQSLKEEFG--PIRNT--NFGSMIS 848
 QY 617 QLEREFVDGERSLAE--SOENAFRKOPAFIOVYL 648
 DB 849 KLETT-----KNKLAELITKADAIKNPSSSKQAL 878

RESULT 16
 Q9KWR3 PRELIMINARY: PRT: 2178 AA.
 AC Q9KWR3: 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 GN Streptococcus hemagglutinin.
 OS HSA.
 OC Streptococcus gordonii.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxId-1302;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DLI:
 RA Takahashi Y., Konishi K., Yoshikawa M.;
 RT "Cloning and characterization of the gene encoding a hemagglutinin of
 Streptococcus gordonii DLI.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB029393; BAA97453.1; -
 DR InterPro: IPR004089; Chmtaxis_transd.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 SQ SEQUENCE 2178 AA; 203507 MW; 75CC27E27F41DB8C CRC64;

Query Match 7.2%; Score 230; DB 2; Length 2178;
 Best Local Similarity 19.9%; Pred. No. 0.33;
 Matches 145; Conservative 131; Mismatches 313; Indels 140; Gaps 24;

QY 10 SGPEVSVSSNOSNMPIINGQI-----ASNSEKSTKASESPSSASSVSS-----SWFL 59
 DB 138 SASSEASANSVSTISISSEFSVSASASLSSSSLSQSSSESASASELSVASTSQSPS 197
 QY 60 SSKAKNLISLRDAIINKNS-----PTDSLQLEASTSTVTRVAKDYD--EAKS----- 109
 DB 198 STTSSTQSSNNSLSSDSNSLNTQSVARNQNRVTRRAVANDTEAPQVKSQDYV 257
 QY 110 -----NFD-----TAKSGLENKTLAEYE-----TKMAD--LMAALODMERLANS----- 147
 DB 258 VYRGSEFEYEAETIDNSGOVNRVYINNEGANGANTYLSFPMWYKYSTENLGRGNATVQNP 317
 QY 148 -----DPSNNHTEEVNNIK-----ALEAQKDTI 171
 DB 318 LTRIFGEVPLINEVEKSYTRYIVAMPDPSGATQMVNANRGLERFVLIVKSNERY 377
 QY 172 DKINKLYVTQNO-----NKSLEVTIKT-----DSAQ 199
 DB 378 DPAESSVTYVNNLSNLSTSEREAVAARAANDNIPPTAKIYVSONGTVITYPPDKSTPT 437
 QY 200 IPA--INSOLEINK--NSADQIITKLERONISYEAVLTNAGEYIKASSEAGIKL--GQALQ 254
 DB 438 IPANRYVKDLQSKNSASQSSSVASQAS-----TSVSASISASMSASVYSTASTS 492
 QY 255 SIYDAGD--QSOQAVALQAOONNSPDNIAATKELIDAETIKVNELKQHTGLTDSPLVKRA 312
 DB 493 ASVSASEASTSASVSASEASTSASVSASKSSSTASVSASE-----SASTSASVSAS 546
 QY 313 EEOISOAKDIOEIKPSGSDIPVGPSSAASAGSAGALKSSNNSGRISLIDVDNEM 372
 DB 547 ESASTSASVSASEASTSASV-----SASTSASTSASVSASEASTSASVS--ASESASTS 600
 QY 373 AATAGGFRSMLEQFVNNPNA--TAKELQAMEQLTAMS--DQLVGADGELPAEIOAID 428
 DB 601 ASVSASEASTSASVSASEASTSASVSASESTASTSASVSASEASTSASVSASEASTS 660
 QY 429 ALAQLAKPSADGLATAMQOVAFAPAAKVGGSAGTAGVQVMYKQLYKTAFTSTSSSYA 488
 DB 661 ASVSASTASTSASVSASTSASVSASTSASTSASTSASTSASTSASTSASTSASTSASTS 720
 QY 489 ALISDGYSAKTYLNSLYSESRSGVSAISQTA-NPALSRVSRSIGIESQGRSADASQRAAE 548
 DB 721 ASVSASTSA--STASVSASTASTSASTSASVSASEASTSASTSASTSASTSASTSAST 779
 QY 549 TIVRDSQTLGDVYSRQVLDLSMTIYSNPQANOEIMOKITASTISAPQPGYAVQNSADSL 608
 DB 780 SASVSASTASTSASVSASES--ASTSASVSASEASTSASTSASTSASTSASTSASESAS 838
 QY 609 DSLQKFAAQ 617
 DB 839 TSASVSASE 847

RESULT 17

0938K3
ID 0938K3 PRELIMINARY: PRT: 1086 AA.
AC 0938K3:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Tail protein.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH1;
RA Ikebe T., Wada A., Inagaki Y., Sugama K., Tanaka D., Suzuki R.,
RA Katsukawa C., Fujinaga Y., Abe Y., Matanabe H.,
RT "Complete sequence of temperate phage PhiNH1.1.".
RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY050245; AAL15083.1;
SQ SEQUENCE 1086 AA; 115592 MW; 8E9329306629F482 CRC64;

Query Match 7.1%; Score 228; DB 2; Length 1086;
Best Local Similarity 21.0%; Pred. No. 0.17; Mismatches 286; Indels 214; Gaps 35;
Matches 165; Conservative 120;

QY 14 SVSSQSSMNPILINGOIASNSETKES-----TKASEA-----SPSSSSVS 54
DB 355 SMSGATGTTIKTYTAVQAVSTMTKADMAKRLSOLGVKASTYTYVMTGALISLSTAATIA 414
QY 55 SWSFLSSAKNALISLDAI-----LNKNSPTDSL-SQLEASTST 93
DB 415 GTAATYATKALVALTGVGVVGAIGALVAVGSLMSLTFKESDETKKLKEGGLVES 474
QY 94 STVTRVAAKDDDEAKSNDDTAKSGLENKTLAEYTKMADLMALQDMERLANSPPSNH 153
DB 475 NKQLE-----DSVEGVOERKKGLESVEKSTAAHQKLADEIKIJAARE-----NKT 520
QY 154 TEEVNIKKALAEOKDTIDKLNKLTVNQNKSLTEVLKTTDSADQIPAINSOLEINK-- 211
DB 521 AGEKRNKLNKIDELNGSIDGLN-LAYDKNSN-----SLSNADQITRSISAMEASTW 572
QY 212 NSADQITKLERONISYEAULTNAGEVTKA--SEAGIKLGAALQSIYDAGD----- 261
DB 573 QTAQONLINTIQQKRESEVSKTLAENALRKKNMEANVSDSVKKEIKIETLTEEAKLKNQ 632
QY 262 ---QQAANVLAQOQNSPDNIAATRE-----LIDAEIKVNEKOE 299
DB 633 TOLQEEYKTSATQQAADAMAAAEESGAROVIAVENMSEARPAIDNMRTKYSLEJET 692
QY 300 HTGLTDSPLVKKAEQISOAQKDIOEIKPS-----GSDIPYVPGSSASASAGSALAKSS 355
DB 693 TTSIFDA-IEQKLTALSVQOMANLEKNRAATEQWATNLEIL-----AQRGVDSILQDL 745
QY 356 NNSG-----RISLLDDVDNEMAAIALQGFMSMIEQFVNNPATAKELQAMEAQLTAMSD 410
DB 746 RRMGEGATQIQVEVDATDAELAPLO-ENFRAATE-----TAK-----NANGS 787
QY 411 QLVGADGLP-----AEIOAIKDA-LAALQKPSADGLATAMGOVAFAPAA 454
DB 788 VLDSAGVMPKPKVGMVTNVSTGLQAELOAANFALQGEIPNGVSQISOGAKSADSAV 847
QY 455 KVGGSAGTACTAGVQMNKOLYKTAFASTSSS-----SYAALSDGVS-----AYTKL 501
DB 848 KMG-----QEVKRSFOGELIGHSFPRVETFEYGHITDGLSNVTNCTSKVMQTM 896
QY 502 NSLYSE-SRSGVQ-----SAISQTAHPALSRSVS-----RSGI-ESQGRSAD 541
DB 897 OSLOAQMSQKQOIVNDRSKSNQITDAFTMSGPMHSHGVANQGLANGIYAGGALAA 956
QY 542 ASQRAETIYVDSQTLGQVYRLOYLDSLMSTIYSNPO-----ANQDET---MQKTLA 591
DB 957 AAGSIAATITATIOSALDIHSPSRVMEDEVGRFI--PGIAGVADIGADRVIDSSMQKLE 1014

QY 592 S--ISKAPQ-----FGYP-----AVONSADSLQKFAALREFEVDGERSLAESQENAFKRP 641
DB 1015 SMTINATPEIASGREGGVAGIANQTTNNSNNSFTLVAVDESQG--NSREKQRLFRFES 1072
QY 642 AFIQQ 646
DB 1073 WYIQQ 1077

RESULT 18
0939N5
ID 0939N5 PRELIMINARY: PRT: 3072 AA.
AC 0939N5:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Platelet binding protein GspB.
GN GSPB.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M99;
RA Bensing B.A., Sullam P.M.;
RT "An accessory sec locus of Streptococcus gordonii is required for export of GspB and for platelet binding.".
RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY028361; AAL13053.1;
DR InterPro: IPR004089; Cmltaxis_transd.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 3072 AA; 285770 MW; 0B148372657CF7F2 CRC64;

Query Match 7.1%; Score 228; DB 2; Length 3072;
Best Local Similarity 20.1%; Pred. No. 0.61; Mismatches 132; Conservative 127; Mismatches 310; Indels 88; Gaps 16;
Matches 132;

QY 10 GSPESVSSNQS-----SNMPIINGOIASNSETKES-----ASPSASSVSMSFL 59
DB 2426 SASESASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASTSASTS 2485
QY 60 SAKAKALISLDAITLNKNSPTDSLQLEASTSTVTRVAAKDDDEAKSNDDTAKSGLE 119
DB 2486 SASESASTS-----ASVSASESASTSAS--VSASTSASTSASVSASESASTSASVSASE 2541
QY 120 NAKTLAEYETKMADLMALQDMERLANSPPSNHTEEVNIIKKALEAKDTIDKLNKLV 179
DB 2542 SASVSASESASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASE 2601
QY 180 LONOKSLTEVLKTTDSADQIPAINSOLEINKNSADQITKLERONISYEAULTNAGEVI 239
DB 2602 SASVSASESASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASE 2647
QY 240 KASSPAGIKLQALQSIYDAGD--QQAANVLAQOQNSPDNIAATKELIDAETKVNELK 297
DB 2648 STSASVSASESASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASE 2705
QY 298 QEHGTLTDSPLVKKAEQISOAQKDIOEIKPSGSDIPVPGSSA-----ASAGSAG 350
DB 2706 ----SASTSASVSASESASTSASVSASE-----SASTSASVSASESASTSASVSASTS 2758
QY 331 ALKSSNNSGRISLLDDVDNEMAAIALQGFMSMIEQFVNNPATAKELQAMEAQLTAMSD 410
DB 2759 ASVSANESASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASE 2818
QY 411 QLVGA-----DGLPAPIOAIKDALQALQKPSADGLATAMGOVAFAPAAVGGGSGTA 464
DB 2819 ASVSASTSASTSASVSANESASTSASVSASESASTSASVSASESASTSASVSASE 2878
QY 465 GTVQMNKOLYKTAFASTSSSSSYAALSDGYSAKTILNLSLSESRGVSQSAISQTAHPAL 524


```

Db      547 TYOAOI-----QAEIAVQ-----VANVEACEKAKTAEQEAKKARDE 583
Qy      418 ELPEIQAIKALAOALQPSADGLATAMGOVA---FAAAKGGGSAGT-ACTGYMNKQ 473
Db      564 AVKAKKEA-EEAKKQAEKAEKTKTATEDANKAKEEAKKASEAKOAEETKADVEEYVA 642
Qy      474 LYKTAFTSSSTSSSYAALSDGYSAYKTLNSLYSESRGVSQSAISOTANPALSRVSRSQI 533
Db      643 V-NVEFESEVKAAKAA-----HHKVPAILDKKKMNAENAKKASAKATEKTTAEAT 695
Qy      534 E-----SOGRSADASRA-----AETIVRDSQTLGDVYSRLQVLDLSMTIVSNP 578
Db      696 KKATEAKTAAGNAQKASENAKAIAADVLAEKASTAQSLKEEAKKL-AADIKKSNVTNDE 754
Qy      579 QANQE-----EIMOKLTASISKAPQFGYPAYQNSAD-SIQKFAQOLEREFVDSERLSAQ 633
Db      755 KAKRKANDAAHQAQSLASAKKEAKTAAQAKGVALEKKKEESAKAVEAKKEMKARD 814
Qy      634 ENAFR-----KOPAFIOQVLYN 650
Db      815 KAAPELLKTKKODVLEQYDVS 835

```

```

RESULT 21
Q9LCJ9 ID Q9LCJ9 PRELIMINARY; PRT: 1795 AA.
AC Q9LCJ9:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE FmtB.
GN FMTB.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1280;
RN RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K5A8;
RX MEDLINE=20346625; PubMed=10896508;
RA Komatsuzawa H., Ohta K., Sugai M., Fujiwara T., Giansmann P.,
RA Berger-Bachl B., Suginka H.;
RT "tn53-mediated insertion inactivation of the fnbB gene encoding a
RT cell wall-associated protein abolishes methicillin resistance in
RT Staphylococcus aureus.";
RL J. Antimicrob. Chemother. 45:421-431(2000).
DR EMBL: AB015223; BAA93430.1;
DR InterPro: IPR000890; Acetate_kin.
DR TIGRFAMs: TIGR01168; VSIRK_signal.1;
DR PROSITE: PS01075; ACETATE_KINASE_1; UNKNOWN_1.
SQ SEQUENCE 1795 AA; 131008 MW; 6CB77CCFCB33D350 CRC64;

```

```

Query Match 7.1%; Score 226; DB 2; Length 1795;
Best Local Similarity 19.1%; Pred. No. 0.38;
Matches 132; Conservative 129; Mismatches 268; Indels 162; Gaps 25;

```

```

Qy      14 SVSSQSSMNPILINGQIASNETKESTKASEKSPASSSVSSWSFLSSAKNALISLRDAI 73
Db      538 AVENMKKDALQOOVNSQVONSHYT-----TASIAEY-----MKLKQOQADPI 577
Qy      74 LNKNSPDSLSOLEASRSTSTVTRVAAKDYDEAKSNFDTAKSGLENKTLAEVETKMAAD 133
Db      578 LNEEDNHVETANRASQADIDGLVTKLQA-----ALIDNOAAIAELDTKAOE 623
Qy      134 LMAALQDMERLANSDP-----SNHTEEVNRIKALEAQKDTIDKLNKLVTLQONQKS 186
Db      624 KVTAAQSGKKTQDEVALVTRKINDKNNAIAEIKQTTSGVTTEKNGIAVLEQD--V 681
Qy      187 LTEVLKTTDSADQIPAT-----NSOLEINKNSAD-----QIKDLE--RQNI 226
Db      682 ITTPVKPQAKODIIQAVTTRKQIKKSNASLQDEKDVANDKIGKIEETKAIKDIDAATNA 741

```

```

Qy      227 SYEAVLTNAGEYIKASSBAGIKLGOALQSIYDAGDOSQAVALQAOONNSPNIATKELI 286
Db      742 QVEAIKTKAINDINOTFTATTAKAALAEFDE-----VQAOIDQALNPDTTNEEV 793
Qy      287 DAEFTKVELKOEHTGLTDSPLVKAERQISOAQKDIOEIKPQSGSDPIVCPGSGSASAG 346
Db      794 AEAIERINAAKV--SGV-----KAIEATTAAQ-DLERVK-----824
Qy      347 SAAGALKSSNNSGRISLLDDVDNEMAIALQGFMSIMEQFN-VNNPATKELQAMEAQL 405
Db      825 -----NEEISKTEINITDSTQTKADA-----INEVQAATARKAQAATVS- 863
Qy      406 TAMSQOLVADGEL-PAEIOAIKDALAOALQPSADGLATAMGOVAFAAKVGSGSAGTA 464
Db      864 NATNEEVAEADAAVEAAQOGIHDIOVYKSKQEVADFTSKYLDKINAIQTO-----AKVK 918
Qy      465 GTQMNKVKQLYKTAFTSSSTSSSYAALSDGYSAYTTLNSLYSESRGVSQSAISQTA----- 520
Db      919 PAADTEVENAVYTRKQEIQNSN-ASTTEEKQAAAYTELDTKKQOEARNTLDAANTNSAVTTA 977
Qy      521 ---NPALSRVSRSQIESOGRSADASORAAE-----TIVRDSQTLGDVYSRLQVLDLSMS 572
Db      978 KDNGIAINQYQAATTKSDAKAEIQAOKASEKRTAIEAMNDSSTTEEQAAAEKVQAVLS 1037
Qy      573 -----TIVSNPQANQOEIMOKLTASISKAPQFGYPAYQNSADSLQKFAQLERE 621
Db      1038 ANADIDNMAANTVDNATTEATTAITPQANVAPQ-----AKQAIKAV-----QAEK 1088
Qy      622 FVDSERLSAESQENAFKOPAFIOQVLYNIA 652
Db      1089 AIDANNQ-STTEEKAAAKQOVTEKTTADAA 1118

```

```

RESULT 22
Q49547 ID Q49547 PRELIMINARY; PRT: 1302 AA.
AC Q49547:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Lmp3 protein.
GN LMP3.
OS Mycoplasma hominis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2098;
RN RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG21;
RX MEDLINE=96213016; PubMed=8631664;
RA Ladefoged S.A., Jensen L.T., Brock B., Birkelund S., Christensen G.;
RT "Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp
RT gene system and identification of gene products.";
RL J. Bacteriol. 178:2775-2784(1996).
DR EMBL: X95601; CAA64858.1;
SQ SEQUENCE 1302 AA; 145799 MW; FB9FD1534014A629 CRC64;

```

```

Query Match 7.0%; Score 225.5; DB 2; Length 1302;
Best Local Similarity 21.6%; Pred. No. 0.26;
Matches 150; Conservative 139; Mismatches 223; Indels 181; Gaps 38;

```

```

Qy      30 IASNETKESTKASASPSASSSVSSWSFLSSAKNALISLRDAIINK-NSPTDLSOLE 88
Db      355 LKSNLD-NEISKAQSLSKDKESMESANDLNTK--LLEYKE-IINKNOEKEAKFNELE 410
Qy      89 ASTSTSTVTRVAAKDY--DEAKS--NPDYAKSGLENKTLAEVET---KMAIDMAALQDM 141
Db      411 Q-----TRKNIENFLLDEVKKNPNVATLVKDLTNADAKKSVYNSNKKDIIIANBAL 463
Qy      142 ERLANSPPSNHNEEVN-NIKKALEAQKDTIDKLNKLVTLQONQKSJTEVLKTTQDSA-- 197
Db      464 IQALADAKKADQOVDEANKSIKEQLNA--LIDKANFTLLPOLNDNDS--EIVKAKESINA 518

```

QY 198 -----DQIPAINSQL-EINKN-----SADQIIKLE--- 222
Db 519 EITNANKAVNONDNAMOSAKSSLDKVTIKQIONLTFEFNKDKAKFKEQETRKIDINLE 578
QY 223 -----RQNIISYEAV---LTAGE---VTKASSAGI-----KLGALQSIIVADGDSQA 266
Db 579 TDDVKNPNPATLVKDLTNAKDKKSVTKSSNKSEITIANDELKQAL-----DKAKYA 631
QY 267 VLAQOONNSPDNIAATKELIDAEFTVNLKQEHGTLDSP-LVKAEQISQAQDIE 325
Db 632 KQIIDEANK-----SIKEGLSDSITNANOLNK---LVSDKDIQAKIELS-----QE 677
QY 326 ITPSGSDIPIVPGSGSASAGSAGALKSSNNGRISLLDDVDNEMAIALOGFRSMTE 385
Db 678 IGSASQELMNNPT-----SMOAKES-----LDAKVTET-----KLE 712
QY 386 ONNVNPNATKELQAMEAQLTAMSDOLVGADGLP-----AEIOAIKALMAQLKQPSA 439
Db 713 TINKDKDVKFEKLETRKDI---DEFINTKNTPNYSTLISELTSKRDSKNSVTNSSK 768
QY 440 DCLATAMGOVAFAAKVGSGSAGTAGTVOANVKQLYKTAFFSSTSSSYAALSDGYSAYK 499
Db 769 SUIETANTLKAQALAKANDKQADLAKSTKEQLNN---SISANTLLAKLTFD----- 819
QY 500 TINSYSESRSQVOSAIQOTANPALSRVSRGIESQSGRADAS---ORAAETIVRD-- 553
Db 820 -KDNITQAKTELEKEV--OKANOAVA-SNNITASMOAKSSLDKAKYTEIKLETFFKDKD 876
QY 554 -----SOTLGVYSRLQVLDLSMSTIVNPOANDEIMQKLTASISKAPQCYPAVONS 607
Db 877 VKFKELEQYRKDI-----DEFINTKNTNP-DYSTLISELTSK-----RDS 915
QY 608 ADSLOKFAALEREFEVGERSLAESQENAFKKQ 640
Db 916 KNSITNSSNKSIDLETANTLKAQALAKANDKQ 948

RESULT 23

Q8RJN9 PRELIMINARY; PRT: 1404 AA.
ID Q8RJN9
AC Q8RJN9
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Variable membrane protein precursor.
GN VMP.
OS Mycoplasma hominis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=132;
RA Boesen T.;
RT "Gene and Protein Structure of the Mycoplasma hominis Vaa adhesin."
RL Thesis (2001), Department of Department of Medical Microbiology and.
DR EMBL; AJ416752; CAC95143.1; -.
KW Signal.
FT SIGNAL
FT CHAIN 1 24 POTENTIAL.
SQ SEQUENCE 1404 AA; 160137 MW; 2C03F66B3473CAB CRC64;

Query Match 7.0%; Score 225; DB 2; Length 1404;
Best Local Similarity 20.4%; Pred. No. 0.3;
Matches 148; Conservative 123; Mismatches 231; Indels 224; Gaps 31;

QY 83 SLSEASTSTSTVTRVAA-----KDYDEAKSNFDTAKSGL-----ENAKTLAEY----- 127
Db 10 SISLIAAASATAIASVCASKNKKFRKQKYNKETEKOELINAKKELKLLLEIKSSDID 69
QY 128 ---ETKMD-----LMAALQDM 141
Db 70 KTNESKVLDSNKIDENSSIEDIQNTQETIEKAIESLTKKINDKNKKHEEDOKIVQAMQEF 129

QY 142 ER-----LANSDDPS-----NNHT-----EYVNNIKKALEAO- 167
Db 130 KKSQKALGDLINSDGQVRVNSNAKOSLQNNNTVNNSSSTEDIIQALSKINEAKKELQSI 189
QY 168 -----KDTIDKLNKLVTLQONKS-LTEVLKTFD--SADQIPAINSQL-EI 209
Db 190 NNARNQKEVEFEKQKQALKINSNEIDNSKRADEFAILKNTNVVVGSDISKIETETKEI 249
QY 210 NK-----NSAQIITKDLERONISYEAVLTNNGEVK--ASSPAGIKLQALQSIYDAD 261
Db 250 EKAIESLTKINEFEKQEKANV--KAVFSKSKQKLDLIDSDGK-----VDSEN 299
QY 262 OSQAAY-LQAOONNSPDNIA-ATKEL---IDAEFTVNLKQEHGTLDSP-LVKAEQI 316
Db 300 ESQVLTKTIDENSSIEDIQNTKDIKAEIESLTNNKINDQKQOKNNLNE--VINKAKELY 357
QY 317 SQAKDIDQIKPSGSDIPIVPGSGSASAGSAGALKSSNNGRISLLDDVDNEMAIA 376
Db 358 KKLVDSDSEIQAKTQLDQEIQKASQVAVASNDTKAINSSKTS--LDAKITDITKK----- 410
QY 377 LOGFRSMIOFNVPNPNATKELQAMEAQLTAMSDOLVGADGLPAPIOAIKALMAQLKQ 436
Db 411 -----LEAFN---ATRK-----LEFTKLETRSNIDKELTPREVKA-----N 443
QY 437 PSADGLATAMGOVAFAAKVGSGSAGTAGTVOAN--VKOLYTAFFS-----TSSSYAA 489
Db 444 PNTTLVNL-ELAKKAKEMVSESSSKSDIVANNLKAQAFQASQKNDADKTSIEKA 502
QY 490 ALSDYSAYKTLNLSYSESRGVOSAI-----QOTANPALSRV 528
Db 503 KLSASLSNAKKLDKNLTDGELQAKAELEVEKANOAITSNNTKEIONSNTSLINKI 562
QY 529 SR-----SGIESQGSADASORAETIVRDSOTLDGVYSRLQVLDLSM----- 571
Db 563 SEYKKNLDFSNKEKEFEKLEKRSATKEFTENNNTNPNYALIDKQAKLDAKKSITK 622
QY 572 ---STIVNPOANDEIMQKLTASISKAPQCYPAVONSADSLQFAALEREFEVGER 627
Db 623 SSNKSIDLIANTQALQALAIKETEKESANSQ--NQAVKTLNLETIKAKELDKNLTDSDG 680
QY 628 SLAESQ 633
Db 681 EITQAK 686

RESULT 24

Q07380 PRELIMINARY; PRT: 1790 AA.
ID Q07380
AC Q07380; P89892;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Hypothetical 206.5 kDa protein YDL058W.
GN USO1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloecher H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z74106; CAA98621.1; -.
DR EMBL; Z74105; CAA98620.1; -.
DR SGD; S0002216; USO1.
DR InterPro; IPR002017; Spectrin.
SQ SEQUENCE 1790 AA; 206450 MW; 90062544F55A52EE CRC64;

Query Match 7.0%; Score 224.5; DB 3; Length 1790;
Best Local Similarity 20.9%; Pred. No. 0.43;
Matches 155; Conservative 130; Mismatches 268; Indels 189; Gaps 32;

```
QY 3 HHHHHHSGPESV-----SSNOSSMN-----PIINGQIASNSE--- 35
Db 969 NNYKDKQAENESLIKAVESSKNSSSIQLSNQKIDMSGOEKNFQIERGSIETKNIQEL 1028
QY 36 -----TKRESTKASEASPSASSSVSSNFSLSAKNAL---ISLDAILINKSSSPDLSIQLE 88
Db 1029 KTIISDLEQFKREELISKSDSSKDEYESQISILKEKLETATJANDEVNKISIELTKTREETLE 1088
QY 89 ASTST-----SYTVRAAKDYDEAKSNFDTAK-SGLENAKTLAEYETKMAADLMAAL 139
Db 1089 AELAAKYNLKNLEETLETLESEKALKVEKNEHEKLEKEKIQLEKEATETKOQLSLNLANIE 1148
QY 140 DMERTLANDSPSNHTEEVNNIKKALE--AQK-----DTIDKLN-KLVTLQONKSLTEVYL 191
Db 1149 SLEK-----EHEDLAAQLKVEEQIANKERQVNEBISQLEINDEITSTQOENESIKK-- 1198
QY 192 KTTDSADQIPAINSS-----QLETKNSADQI---IKDLEKONISYEAVLTNAGEVIKASSE 244
Db 1199 KNDLEGEVKAAMKSTSEESQSNLKSEIDALNLQIKELKKRNETNEASL---ESIKSVES 1255
QY 245 AGIKILGO-----ALOSIVDAGDQQAVALQAOONNSPDNIATKELIDAEFT 291
Db 1256 ETWKIKELQDECNFEKEVSELEDKLASEDKNSKYLEIQKESE---KIKEELDAKTT 1310
QY 292 KYNELKQOEHTGL-----TDSPLVKKAEEQISOAKDIOEIKPSGSDIP 335
Db 1311 ELKIQLEKTTNLSKAKKESSELSRLKTSSEERKNAEEQLEKLEKNIQ-IRKQAE--- 1366
QY 336 VPPSGSASAGSAGALKSSNSGRISLLDDV-----DNEMAAIALOGFRSMIEQFNVA 390
Db 1367 --KERKLNEGSSST---ITQYSEKINTLEDELIRLQNEENELKAKKIDMTRESELEKVSLS 1421
QY 391 NPA-----TAKELQ---AMEAQUTAMSDOYVAGDGLPAEIQAIKALQALQOPSA 439
Db 1422 NELLEEKONTIKISLDDELSTKYDKITTRNDEKLLSTERNNKRDLEELKQLRAAQ--- 1477
QY 440 DELATAMQVAPAAKAVGGSAGTACTVQMNVKQLYKTAFFSSSTSSSYAALSIDGYSAYK 499
Db 1478 -----SKAVEEG-----LKKL-----EESSEKKAEELEKSEKEMMK 1508
QY 500 TINSLSYSESRSGVSAISQTPANPALSRVSRSGISQGRSADASQRAEFTIVRDSQ-TLG 558
Db 1509 KLESTIESNETELKS-----SMETIRKSDKLEQSKSAEEDIKNLQHEKS 1554
QY 559 DYYSRJ-----QYULDSLMS---TIVSNPQANOELMOKLTLSISKAPPGYPAVONSADSIQ 612
Db 1555 DLISRINSEKDIIEELKSKLRLEAKSGSELETVQELNNAQEK---IRINAENTV--LK 1609
QY 613 KFAAOLEREFEVDEGRSLAESOE 634
Db 1610 SKLEDIERELKDKQAEIKSNQE 1631

RESULT 25
Q9X7M2 PRELIMINARY: PRT: 1327 AA.
AC Q9X7M2:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE CTOREF1365, partial (Fragment).
OS Staphylococcus aureus.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacilliales;
Staphylococcus.
OX NCBI_TaxId=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COL:
RX MEDLINE-99265121: PubMed=10332717;
RA "Mrp-a new auxiliary gene essential for optimal expression of
RT methicillin resistance in Staphylococcus aureus.";
RL Microb. Drug Resist. 5:9-18(1999).
```

```
DR EMBL: Y09928; CAA1062.1; -.
FT NON_TER 1
FT NON_TER 1327
SQ SEQUENCE 1327 AA; 140517 MW; B90F2085E800586D CRC64;

Query Match
Best local similarity 19.6%; Pred. No. 0.37;
Matches 136; Conservative 114; Mismatches 275; Indels 170; Gaps 26;

QY 14 SVSSNSSNNPIINOIAENSETKESTKASEASPSASSSVSSNFSLSAKNALISLRDAI 73
Db 496 AVEMNKDALQOQVNSQVDSNHYT-----TASIAEY-----NKLKQOADPT 535
QY 74 LKNSSPDLSOLEASTSTSYTVRAAKDYDEAKSNFDTAKSGLENAKTLAEYETKMA 133
Db 536 LN-----EDANHVTANRASQADIDGLVTKLQAL--IDNQAAIAIELDTRKAE 581
QY 134 LMAALQDMERTLANDSP-----SNHTEEVNNIKKALEAQKDTIDKLNKLVTLQONKS 186
Db 582 KYTAAQOSKKVTQDEVAALVTKINNDKNNAIAEINKQTAAQVTEBKDNGIAVLQDD--V 639
QY 187 LREVLKTTDSADQIPAI-----NSQLETKNSAD-----QIIKDLE--RONI 226
Db 640 ITPTVKPOAKODIQAIVTTRKQOIKKSNASLDQEDKDVANDKIGKLETKAIKDIDATTVA 699
QY 227 SYEAVLTNAGEVTKASSEAGIKLGOALOSIVDAGDQQAVALQAOONNSPDNIATKELI 286
Db 700 QVEAIFTKAINDNOTTPATTKAKAALFEDE-----VQAOIQDQAPLPDPTTNEEV 751
QY 287 DAAETKVNELKQOEHTGLTDSPLVKKAEEQISOAKDIOEIKPSGSDIPVPGSGSAAAG 346
Db 752 AEATERINAKV--SGV-----KAIEATTTAQ-DLEHYK----- 782
QY 347 SAAGALKSSNSGRISLLDDVDNEMAIALOGFRSMIEQFN-VNPNATKELQAMEADL 405
Db 783 -----NBEISKIENIDISTQTKMDA-----YNEVKQAATARKQONTVS- 821
QY 406 TAMSQDQVAGDEL-PAEIQAIKDALQALQOPSADGLPANGQVAPAAKAVGGSAGTA 464
Db 822 NNTNEEVAEDADAQAAQOGGLHDIOVYKSKQEVADTKSKVLKINAQTQ-----AKYK 876
QY 465 GTVQMNVKQLYKTAFFSSTSSSYAALSIDGYSAYKTLNLSYSESRGVO----- 513
Db 877 PAADTEVENAYNTRKQEIQNSN-ASTTEKQAAYTELDTRKKQEARFNLDAAANTNSDVTTA 935
QY 514 -----SAISQT-----ANPALSRVSRSGISQGRSADASQRAEFTIVRSOTL 557
Db 936 KNSIAAINQOVAATTKSKDAKAEIQAOKASERKTAIEAMNDSTTEQOAA---KDKVQ 991
QY 558 GDVYSRLQYULDSLMSFTIVSNPQANOELMOKLTLSISKAPPGYPAVONSADSIQFAAO 617
Db 992 AVYTANADIDNAANDVNNAKTTNATTAITPPDANKVP-----AKQAIADKV-----Q 1042
QY 618 LREFEVDEGRSLAESQENAFRRQAPFIQOVLVNI 652
Db 1043 AQETALIDGNGG-STTEKKAQAQOVQTEKTTADAA 1076

RESULT 26
Q9RL69 PRELIMINARY: PRT: 2478 AA.
AC Q9RL69:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE MRP protein.
DE MRP protein.
OS Staphylococcus aureus.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacilliales;
Staphylococcus.
OX NCBI_TaxId=1280;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN-COL;
 RX MEDLINE-97302526; PubMed-9158773;
 RA Wu S., de Lencastre H., Sali A., Tomasz A.;
 RT "A phospholuciferase-like gene essential for the optimal expression
 of methicillin resistance in *Staphylococcus aureus*: molecular cloning
 and DNA sequencing.";
 RL Microb. Drug Resist. 2:277-286(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COL;
 RX MEDLINE-97431478; PubMed-9286983;
 RA Jolly L., Wu S.W., Van Heijenoort J., de Lencastre H.,
 Wenglin-Iecreux D., Tomasz A.;
 RT "The femR35 gene from *Staphylococcus aureus*, the interruption of
 which results in reduced methicillin resistance, encodes a
 phospholuciferase mutase.";
 RL J. Bacteriol. 179:5321-5325(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COL;
 RX MEDLINE-99265121; PubMed-10332717;
 RA Wu S., de Lencastre H.;
 RT "Mip-a new auxiliary gene essential for optimal expression of
 methicillin resistance in *Staphylococcus aureus*.";
 RL Microb. Drug Resist. 5:9-18(1999).
 DR EMBL: Y09927; CAB5329.1; -
 DR TIGRFS: TIGR01167; LPTXG.anchor: 1.
 DR TIGRFS: TIGR01168; YSIRK.signal: 1.
 SQ SEQUENCE 2478 AA; 263031 MW; 6B9859A02D023C74 CRC64;

Query Match 6.9%; Score 222; DB 2; Length 2478;
 Best Local Similarity 19.6%; Pred. No. 0.81;
 Matches 136; Conservative 114; Mismatches 275; Indels 170; Gaps 26;

QY 14 SVSSNOSMNPITNGOIASNFKSTKASEASPSASSSVSSPFLSAAKNALISLRLDAI 73
 DB 535 AVEMNDALQOOVNSODNSHTY-----TASTAET-----NKLKQOQADTI 574
 QY 74 LKNSSPDLSOLEASTSTSTVRYAAKDYDEAKSNFTPAKSGLENAKTLAEYETKMA 133
 DB 575 LN-----EDANHHKKTANRASQADIDGLVTRKLAAL--IDNQAIAELDLTKRAE 620
 QY 134 LMAALODMERLANSDP-----SNHTEVNNIKKALEKQKDTIDKLKLVTLONONKS 186
 DB 621 KVTAAOQSKKVTODEVAALVTKINNKNNAIAEINKQTTAQGVTEKDNIAVLEOD--V 678
 QY 187 LFEVLKTTDSADQIPAI-----NSOLEINKNSAD-----QIILKLE--RQNI 226
 DB 679 ITPTVPRQAKODITIOAVTTTKKQIKKSMASLQDEKDVANDKIGKIEKTKIDIDATTTNA 738
 QY 227 SYEAVLTNAGEVIRKASEAGIKLQALQSIYDAGDQSOAAVLQAOQNNSPDNIAATKELI 286
 DB 739 QVAETIKATINDIOTTPTATKAALAEFDE-----VVOAQIDQALINDTINEEV 790
 QY 287 DAAETKYNELKQHTGLTDSPLVYKKAEEQISQAKODIQETKPSGSDIPIVPGSGSAASAG 346
 DB 791 AEAIERINAKV--SGV-----KAIEATTTAQ--DLERVK-----821
 QY 347 SAAGALKSSNNSGRISILLDDVDNEMAAIALQFRSMIEFQ--VNNPATAKELOAMEAQL 405
 DB 822 -----NEISKIENTISTOTKMDA-----YNEKKQAATARKKQONATVS- 860
 QY 406 TAMSOLVYGDEL-PAEIOAIKDALAQLKOPADGLATAMGOVAFAAKVGGSAGTA 464
 DB 861 NAINEEVAEADAVIDAQAQKGLHDIOYVRSKQEVADTKSKVLKINAIQOQ-----AKV 915
 QY 465 GTVQNMVKQLYKTAFFSTSSSSYAALSDGYSAVKTLNLSYSESRSQV-----513
 DB 916 PAADTEVENAYNTRKQIQSN--ASTEEKQAAYTELDTKQEARITLDAANTNSDVTTA 974
 QY 514 -----SAISQT-----ANPALSRVSYSIGTESQSRASDASRAAETIVRSQTL 557
 DB 975 KONSIAAINVOAATTKKSDAKAEIAOKASERTTAIEAMNDSTTEEOAA-----KDKVDO 1030

QY 558 GDVYSRLQVLDLSMTITVSNPQANOELMOKLTASISAKPOFGYPAYONSADSLQFAAQ 617
 DB 1031 AVYTANADIDNMAANDVDNNAKTNEATITATIPDANVPR-----AAQKAIADKV-----Q 1081
 QY 618 LEREYVDGERSLAESOEANFRKQPAFIQOVLVNI 652
 DB 1082 AQETALDGNNG--STTEKAAARQOVOTEKTTADAA 1115

RESULT 27
 ID 015738 PRELIMINARY; PRT; 924 AA.
 AC 015738;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
 DT 01-JAN-1998 (TREMBLrel. 05, last annotation update)
 DE zipa (Fragment).
 GN zipa.
 OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RA Loomis W.F., Trautfar N.;
 RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF01980; AAB70839.1; -
 FT NON_TER 924
 SQ SEQUENCE 924 AA; 105060 MW; 63921CCA596AB406 CRC64;

Query Match 6.9%; Score 221.5; DB 5; Length 924;
 Best Local Similarity 21.1%; Pred. No. 0.25;
 Matches 148; Conservative 122; Mismatches 282; Indels 149; Gaps 29;

QY 12 PESVSSNOSMNPIT--INQIASNFKSTKASEASPSASSSVSS-----FLSAAKN 64
 DB 278 PESNSPINOSTNDLQOYNTLNGNNNTTTNKKKEITELQSLNISAKEIMVQYQ 337
 QY 65 ALISLRLDAI---LNKNSPDTLSOLEASTSTSTVTRV-----AAKDYDEAKSNFTPAKSG 117
 DB 338 QIVTLOQOVSFTEKYDDLTLNLSQKE--TKIGELTRATNGFTTKETILRSYEDKKRT 395
 QY 118 LENAKTLAEYE-----TKMADLMAALODMERLANSDPNNHTEVNNIKKA 163
 DB 396 AELLERLEWEYKKNKNTDEKDFQIEKIVDQLEAKQSEQOTT---TNNIQNEISOLKQ 451
 QY 164 LEAKQKT-----IDKLKLVTLONONKSILTEVAKTDS-----ADQIPAINSOLEINKNS 213
 DB 452 LASNSTESQALQSKITELSQLQSEFEKLQNOQSKDSELETETSKROSALLEQOSEDSSOS 511
 QY 214 ADQIILKLE-----RONISYEAVLTNAGEVIRKASE-----AGIKLQALQS 255
 DB 512 KDEKLSVLEINLOQTLOQLOSKQDELONVKSQLEQSESESKDQKLSVEILTLQOTLQ 571
 QY 256 IYDAGQSOQAALQAOQNNSPDNIATKELIDAEETKYNELKQHTGLTDSPLVYKKAEEQ 315
 DB 572 LDVKKSOLEQ--QSEHNESKD-----EKLSIEIENLQOQLOS---PDSELSRK--DEQ 617
 QY 316 ISQAQKDIOEIKPSGSDIPIVPGSGSAASGSAAGALKSSNNSGRISILLDDV---DNEM 372
 DB 618 LKCLESELISVKKQOLSS-----QSSNTDSELSVKKDQLSKDSEL 657
 QY 373 AA-----IALQFRSMIEQFVNNPATAKELOAMEAQLTAMSODLVGADGE---LP 420
 DB 658 KSKDEQLSKDSQKSTIESDQSVKQDLSKQDELQSTKQDQLSKDQSLNKTQIKSTIE 717
 QY 421 AETIOAIKDALA-----QALK-----QPSADGLATAMGOVAFAAKVGGSAGTA 464
 DB 718 SDLQSVKQDQLSKDQELQSTKQDQLSKDQELQSTKQDQLSKDQELQSTKQDQLSKDQEL 774
 QY 465 GTVQNMVKQLYKTAFFSTSSSSYAALSDGYSAVKTLNLSYSESRSQVSAISQITNPAL 524

```

Db 775 TTDLASDK---TELSTKDKQ-----LSSKDELSQISKDQSTKDELSQSKDQLS----- 822
QY 525 SPSVSRSGIESGGRSADASQRAAEFTV--RDS---QTLGDVYSRLQVLDLSMSTVSNPQ 579
Db 823 SKDSELSQSTKQDLSKSDSLQSVKQDQLSKSDSLQSTKQDLSKQDQLSQVADQLSTKQD 882
QY 580 ANQEIWMKLTASISKAPQFGYPAYONSADSLQ-KFAAOLE 619
Db 883 ELQ-QITSKQSEQDSKVSQIQDQDLKNAKNAEFLSYTFEKQTE 922

RESULT 28
Q96200 PRELIMINARY: PRT: 1627 AA.
AC 096200:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Axoneme-associated protein GASP-180.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
CX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RA Elmendorf H.G., Rohrer S.C., La Vigne E.A., Nash T.E.;
RT "Novel Axoneme-Associated Proteins in Giardia lamblia.";
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF400249; AK91740.1; .
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 5.
DR PROSITE; PS00089; ANK_REPEAT; 2.
DR PROSITE; PS02997; ANK_REPEAT_REGION; 2.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
KW ANK repeat; Repeat.
SQ SEQUENCE 1627 AA; 179270 MW; 021ED9763907DCC CRC64;

Query Match 6.9%; Score 220.5; DB 5; Length 1627;
Best Local Similarity 20.9%; Pred. No. 0.55; Indels 177; Gaps 32;
Matches 152; Conservative 131; Mismatches 268;

QY 16 SSNQSMPPIINGQIAINSSEKSTKASPSASSSVSSWSFLSSAKNALISLR----- 70
Db 262 SSTVGQNVSLHGQY-KNNEENITLRIEISLQSTAKDKAIIIDLKRLDMQEMLSA 320
QY 71 -----DAITKNSSPDLSLSQLEAS-----TSTSTVTRVA--KDYD-----AKS 109
Db 321 GSGGSDVGLKDKIRALQSELAEKERDIRALNERLTAVPSPGDAALLADYEKKISALRE 380
QY 110 NEDTAKSGLENAKTLAE-----YETKADLM-----AALQDMERLANSDPNNHT----- 154
Db 381 ELEAREESDSGLRPVEKTRAFEDLAGRLSMDEKDAITAELEQAAODEAVATQIIE 440
QY 155 EEVNNIKKALEAKQDPTIDKLKLVTLQNONKSLTEVLTAKTTSDADQIPAI-----NSQL----- 207
Db 441 EEVESLRQALLA-----KDGELEALQSR-----LEETLANGGEADPDATIAAKNSIARL 491
QY 208 -----EINKNSA--DOITKLEKONISYEAVLTNA-----GEVITAKSSEKICIL 249
Db 492 ODLVKRAESQRRSSAEGDKRVAQLQELRLLADAGGDDERKVELNEQIADQFEL 551
QY 250 GOALQSIYAGQSOAQLQAOONNSPDNIATKLELIDAEETKVNKLEQHEHGLDPLV 309
Db 552 ANAKQDIDERDEIF--LNRDRIQEMKNSALQERVDLE-----ADATRG 596
QY 310 KRAEQDISQAKDQIEIKPSGSDIPVPGSGAASAGALKSSNNSGRISLLLDVY- 368
Db 597 ADAAEVLARIEEQGVK-----ELSGRSRQAPADETAIRLEE-----KIRALDDEIE 645
QY 369 --DNEMAAIALQGFRRSMITQFVNNNPATAKELQAMEAQITAMSQVLVGADGELPATQAI 426
Db 646 ARDNOIAEL-----KELVD-----GAPQAPVEADPAQTLALEE-----NGRLKELQAL 690

```

```

QY 427 KDALAQAOLKPSADGLATAGOVAFAAKVG-----GSAGTAGTQVMNKKQLYTFASS 481
Db 691 NDAL-FALRKSSADEASGLRGVAHLNKRVSQLEKLAARASGDAS--DYKLVLEQBEL 748
QY 482 TSSSSYAALSDGYS-AVYKTLNSLYSESRSGVQSAISQTPAN-ALSRSYSR----- 530
Db 749 EDAQEQLLSLKRDYDCAVAMEDMRLEPQKPVGSTVYEEPGASSEDLDRLKEELDALR 808
QY 531 -----SCIESQGRS-----ADASQRAAEFTVND-----SQTIGDVYSRLQVLD 568
Db 809 EEVOYVLENEELVYHMGONRQKDELNRLNGALDAKESLADLKAQDSTVPQDARKILIE 868
QY 569 SLMSTVSNPQANOBEI--MOKLTASISKAPQFGYPAYONSADSLQKFAAOLEREVDGE 626
Db 869 DEIADLRGTVAARDDAIRGLEKTAIRLAE-----LEQLAADKGEVADKE 913
QY 627 RSLAESOE 634
Db 914 HSLRLLED 921

RESULT 29
Q9VM67 PRELIMINARY: PRT: 1833 AA.
AC Q9VM67: Q9VM66:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG18304 protein.
GN CG18304 OR CG13774.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazef J.R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agdayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Switskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

```

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003615; AAF52458.2;
 DR FlyBase: FBgn0031869; CG18304.
 DR InterPro: IPR002017; Spectrin.
 KW Hypothetical protein.
 SQ SEQUENCE 1833 AA; 206010 MW; ED1948482573584E CRC64;

Query Match 6.9%; Score 219.5; DB 5; Length 1833;
 Best Local Similarity 21.6%; Pred. No. 0.7;
 Matches 145; Conservative 105; Mismatches 265; Indels 155; Gaps 26;

QY 16 SSNOSNMPIIINQIANSSETEKSTKAS-EASPSASSSVSSWSFLSAKKNALISLRDAIL 74
 DB 250 SSNNS-----TSLSNSTSTASNEVKYVTSTSSSTSSSVRRKREADVASKEL 299
 QY 75 NKNSSPTDSLSLEASTSTSTVTRVAAKYDE--AKSNFTASGLENAKTIAE-YETKM 131
 DB 300 KROTVPASTISH--SNSTSTASTASKSODINGMOEQMKALTELEMTKTRAKERERK 356
 QY 132 ADMAALQDMERLANSPPSNHT--EEVNNIKKALEAKQDTID-----KLN-KLVITIQ 181
 DB 357 SDLL-----LRLASMDTASNRTAASEBALNLQKLNEMKQDLRVYEDKKRLRMKELE 411
 QY 182 NQNKSLTEVLTQTSADQIPAINSOLEINKNSADQIIKLERONISYEAVLTNAGEVIKA 241
 DB 412 NKS-----ESELRRKLQAAEQICEELMEKNQSKKLTIL----- 446
 QY 242 SSAGIKLQALQSIYD-AGDQQAAYLQAOQNNSPDNIAATKEL-----IDAETKVN 294
 DB 447 -----LQAEDEVDQDFRDEEVAKRTSLQKLEKATNCRILSEFLKKSRIE 495
 QY 295 ELKQEHGLTDSL--VKKAEQI-----SOAKDQIEIK-PSGSDIPITVPGSGSAA 343
 DB 496 TLEQERQSSNALSKIKKLEELRFSNELTKLQAEAEELNPGKKAPMLGVLGKST 555
 QY 344 SAGSAAGALKSSNSGRISLLDDVDNEMAALQGFRRMIEQFNVNPA-TAKELQAME 402
 DB 556 SA-DAKFTRESLTRGG-----SQEDPQHLORELQDSI 586
 QY 403 AQLTAMSDQLVGADGELPAETQAIKDALQALQPSADGLATMGQVAAKVGSGSAG 462
 DB 587 ERTDTLKDQKFAEEL--QRLRDERKRVFSCGTQTEVLEVAFAFR-----GTQ 636
 QY 463 TACTVOMNKOIKTAFSSSTSSSYAALSDGYSAYKTLNLSSESQVQSAISQTAMP 522
 DB 637 TVATVOSD-----STSEVNLVTSNVAVTQD-----FEVPRNVSIETETWSPFAGLEPP 688
 QY 523 ALSRSVRSQIE-----SQGRSADASQRAETTVRSQTLGDVYSRL 564
 DB 689 SSSSRVGGSGRKLSPTPHRHLAEVHADREGIS-DEDDPALRLILLNEQEAISLRL 747
 QY 565 QVLDLSMTITVSPQANQOEIIMOKLTASISKAPQFGYPAVQNSADSLQFAOLEFEVD 624
 DB 748 KYEDLEKENAESKKYV-RELQAKLRQDSSNGSKSLSLSTGSSAAEKVKVLTINEELVQ 805
 QY 625 GERSLAESOE 634
 DB 806 LRRTLEKEQ 815

RESULT 30
 ID 007569 PRELIMINARY; PRT: 2139 AA.
 AC 007569; 002504;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-JUN-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Myosin heavy chain.
 GN MHCA.
 OS Entamoeba histolytica.

OC Eukaryota; Entamoebidae; Entamoeba.
 OX NCBI_TaxID=5759;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=HMI:IMSS;
 RX MEDLINE=93295430; PubMed=8515774;
 RA Raymond-Denise A., Sansonetti F., Guillen N.;
 RT Identification and characterization of a myosin heavy chain gene
 RT (mhca) from the human parasitic pathogen *Entamoeba histolytica*.
 RL Mol. Biochem. Parasitol. 59:123-131(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HMI:IMSS;
 RA Guillen N.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: L03534; AAB48065.1; -
 DR HSSP: P08799; 1MND.
 DR InterPro: IPR001609; myosin_head.
 DR InterPro: IPR000048; IQ_region.
 DR Pfam: PF00612; IQ_2.
 DR Pfam: PF00063; myosin_head; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.
 DR SMART: SM00242; MYSC; 1.
 SQ SEQUENCE 2139 AA; 245225 MW; C68307341DB51DD1 CRC64;

Query Match 6.9%; Score 219.5; DB 5; Length 2139;
 Best Local Similarity 20.9%; Pred. No. 0.85;
 Matches 151; Conservative 120; Mismatches 272; Indels 179; Gaps 32;

QY 13 ESYVSSNOSNMPIIINQIANSSETEKSTKASEASPSASSSVSSWSF-----LSAKNAL 66
 DB 994 ESIDEKEDEITK-LKGDIKLLEEKEDLQDRADVATKQDLIAKKINKITIECEDAKDEL 1052
 QY 67 ISLRDAI--LNKNSPTDSLSLE--ASTSTVTRVAAKYDAKSNFTAKSGLN 120
 DB 1053 AKLEQLEDEENKNDLTJELQTLKLGFTESLAAQVAAT--KASDERDTLSQNLN 1110
 QY 121 -----AKTLAEVETKMAADLMAALQDMERLANSPPSNHTFEVNNIKKALEAKQDTID 172
 DB 1111 EKLTTLNLTNRKADLEKRTISGLKQDYEDLE--DDKKNIEDDLNAQRIKLEDEIT 1165
 QY 173 K-----LNKLVTLOQNKSLTEVLTQTSAD-----QIPAINSQL--ET 209
 DB 1166 KGADEVQYLQKQKEEYESQIAKQOEKEAIGNDVKNKKEKTIKEKELEIOSLQKLEDETEV 1225
 QY 210 NKSADQIIKDL-----ERQNTSYEAVLTNAGEVIKASEAETIKLQALQSIYDVG 260
 DB 1226 EKEDAKKKKEIKEKKALQOEKEVNESS--KNSTEKKKKLELDNLKDTQKKLDMDTAD 1282
 QY 261 DQ--SQAAVILQAOQNNSPDN--IAATKELID--AAETKVNELKQEHGTGLT----- 304
 DB 1283 NEKLAKKAKDLEQNLNBYQDNHKKAVADAEILLNKKRAQSDKELNSLAKALEALTRAKSYV 1342
 QY 305 -----DSPLVKKK-AEQISOAKDQIEIKPSGSDIPIVPGSASASAGSAGALKSSNNS 358
 DB 1343 ESKNKDSENEKALSEIIOANEKLNIO--ADL-----RKATADLOEANEKKAEVEA 1393
 QY 359 GRISLLD-----DVND-----EMAAITLQGFRRMIEQFNVNNTATAELEQ 400
 DB 1394 QRDKLIVADNKKMTKLTLEITKARDEENTYVEENEYKLLKREADELEFANENLDEKKDRNN 1453
 QY 401 MEAQLTAMSDQLVGADGELPAETQAIKDALQALQPSADGLATAMQVAAKVGSGS 460
 DB 1454 KEQVAKKLEGELEKTEKDKLNAAI-AEKDSIFPAKKQSDAD----- 1492
 QY 461 AGTAGTVQANVKQLYTA-----FSSTSSSYAALSDGYSAYKTLNLSSESRSQVQ-- 513
 DB 1493 -----LEELNKTEVEHDEEVVAKLTQITKLRDQNSAEELNELRSKADKKKKRI 1542
 QY 514 SALSQTAMPALSRVSRSGLSGRSADASQRAAETIVDSQTLGVYSRLQVLDLSMT 573
 DB 1543 SELEQVNELESRPVGTG-----NADEN--EIKIRDAQ-IADLNKALE--MKG 1585


```

QY 574 IVSNP-QANOEIMOK---LTASISAKPOGYPAYONSADSLQKFAOLEREFVDCERSL 629
DB 1586 VONNOLOATYNKELKAKNDLTKIE-----ITENEMKLENAKKRLEQDKDEADKAV 1637
QY 630 AE 631
DB 1638 SE 1639

RESULT 31
08SY55 PRELIMINARY: PRT: 1514 AA.
AC 08SY55:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE GH09355p.
GN CG6004.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceiniker S.;
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY015323; AAL68190.1;
SQ SEQUENCE 1514 AA; 157140 MW; 1FFC4B0664105AD5 CRC64;

Query Match
Best Local Similarity 6.8%; Score 219; DB 5; Length 1514;
Matches 132; Conservative 125; Mismatches 321; Indels 114; Gaps 17;

QY 11 GPESVSSNOSMNPILINGOIASNSETKESTKASEASPSASSVSWSPFLSAKNALISLR 70
DB 414 GPLSTESSSTREATNESSSTSSODSTTQESSSTSEGPLSTESSSTREATNESSSTE---SSQ 469
QY 71 DALINKNSPTDLSOLEASTSTSTVTRVAANDYAKSNFTAKSGLENAKTLAEYETK 130
DB 470 DSTQESSSTSTGPLSTESSSTREATNESSSTSSODSTTQESSSSSEGPLSTESSSTE-ATN 528
QY 131 MADLMAALQDMERLANSDSNHTEVNNIKKALEAQDXTIDKLNKLVTLQONKSLTEV 190
DB 529 ESSSTESSOD---STTQSSSTSTESPLSTEPSTEANESSSTSSODSTTQESSSTEDP 584
QY 191 LKTTDSADQIPAINSOLEINKNSADQIILDKERONISYEAVLITNAGEVITKASSEAGIKIG 250
DB 585 LSTESSTE-----ANNESSSTSSODSTTQ-----ESSSTSEPLSTESSSTESNS 631
QY 251 QALQSTIVDAGDSQAQAVLDAQOONNSPDNIAATKELIDAAETKYVNEIKQHTGTLDSPLYK 310
DB 632 SSTESSODSTTQSSSTSTESPLSTEPSTEANES---STESSODSTTQESSSTSEGPLST 688
QY 311 KAEQISOJ-----AKQDIOEIKPSSGSDPIVPGSGAASAGSAGACALSS----- 355
DB 689 EPSTEANESSSTSSODSTTQESSSSSEGPLSTESSSTEANESSSTSSODSTTQESSST 748
QY 356 -----NNSGRISILLDDYDNEMAAIALQGRSMIEQFNVNPAATAKELQAMEA 403
DB 749 ESPLSTEPSTEANESSSTSSODSTTQESS--STEGPLSTEPSTEANESSSTSSODSTT 807
QY 404 QLTAMSDQVLGADGELPAE-----IOAIKALQALQOPASDGLATANGOVAFPA 453
DB 808 QESSSS-----SEGPLSTESSSTEANESSSTSSODSTTQESSSTEDPLST----- 853

```

```

QY 454 AKVGGSGACTAGTVOMNKKLYTAFTSSSSSYAAALS-----DGYSAVKTINSL 504
DB 854 -----ESSTEATYESSSTSSODSTTQESSSTSEGPLSTESSSTESSSTSSODST 907
QY 505 YSESRGVSQSAI-----SOTANPALSS-----RSVSRGIESQGRSADASQRAAET 549
DB 908 TQESSSTSEPLSTEPSTEANESSSTSSODSTTQESSSTSEGPLSTESSSTEANESSSTE 967
QY 550 IYRDSOTLQDVYSRLQVLDLSMTSTIVSNQANOEIMOKLT---ASISAKPOGYPAYQ- 605
DB 968 SSODSTTQESSSTSEGPLSTESSSTEGNSSTSSODSTTQESSSTSEPLSTEPSTE 1027
QY 606 ---NSADSLQKFAOLEREFVDCERLAEOS 634
DB 1028 NESSSTSSODSTTQESSSTSEGPLSTESSSTE 1059

RESULT 32
09ZFF9 PRELIMINARY: PRT: 2570 AA.
AC 09ZFF9:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Fimbriae-associated protein Pap1.
GN Pap1.
OS Streptococcus parasanguis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FW213;
RA MEDLINE=20063153; PubMed=10594831;
RA Wu H., Fives-Taylor P.M.;
RT Identification of dipeptide repeats and a cell wall sorting signal in
RT the fimbriae-associated adhesin, fap1, of Streptococcus parasanguis.";
RL Mol. Microbiol. 34:1070-1081(1999).
DR EMBL: AF100426; AAC79868.1;
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR001313; Pumlilio/Puf.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR TIGRfams: TIGR01167; LpXTG_anchor.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING;
SQ SEQUENCE 2570 AA; 263107 MW; B90ED356754409A8 CRC64;

Query Match
Best Local Similarity 6.8%; Score 218.5; DB 2; Length 2570;
Matches 140; Conservative 151; Mismatches 288; Indels 133; Gaps 28;

QY 13 ESYSSNOSMNPILINGOIASNSETKESTKASE-----ASPSASSVSWSPFLSAKNALI 67
DB 70 ETVLAKETTLTTTDAANEVLSENFDESEKAERKISLSQESASSEVS--ESISSEVSSEV 127
QY 68 SLRDATLNK-----NSSPTDSLQ--LEASTSTSTVTRVAANDYAKSNFTAKSG 117
DB 128 STSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESV 187
QY 118 LENAQTALAEYETKMADLMAALQDM--ERLANSNHNTEVNNIKKALEAQDXTIDKLN 175
DB 188 EKKODSVRENLDKMISEAEVLNDMAARKLITLD---AEQOLELMKSLVATQSOLEATK 242
QY 176 KLVTLQONKSLTEVILKTTDSADQIPAINSOLEINKNSADQIILDKERONISYEAVLITNA 235
DB 243 NLIGDPNATVADQLATYTTLGNNTQALGNELKLNPG--QIYAVLNTTEASRAATL--- 297
QY 236 GEVIRKASEAGIKLQALQOISYDAGDSQAQAVLDAQOONNSPDNIAATKELIDAAETKYV- 294
DB 298 -----RSTTTGKTFTTIDFSGNGQIYYWA--GGANANLKPISISAVIYSATGKISW 350
QY 295 ELKQHTGTLDSPLYKKAEOIS---QAOKDIOEIKPSSGSDI-----PIVPGSGSA- 343
DB 351 TVEYDPTTLKSPALKTKITTYGIVIDTSSDSKLSPTFVNLIDGAATNVTNFTYNGSGKG 410

```


01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Extracellular matrix binding protein (Fragment).
GN EMB.
OS Abiotrophia defectiva.
OC Bacteria: Firmicutes: Bacillus/Clostridium group; Lactobacillales;
OC Aerococcaceae; Abiotrophia.
OX NCBI_TaxID=46123;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NS-47;
RX MEDLINE=99081722; PubMed=9864195;
RA Manganello R., van de Rijn I.;
RT "Characterization of emb, a gene encoding the major adhesin of
RT Streptococcus defectivus.";
RL Infect. Immun. 67:50-56(1999).
DR EMBL: AF067776; AAD0320.1; -
DR TIGRfams: TIGR01168; YSIRK_signal. 1.
FT NON_TER 2055
SQ SEQUENCE 2055 AA; 215642 MW; 9699C11DDE93E2FD CRC64;

Query Match 6.8%; Score 216.5; DB 2; Length 2055;
Best Local Similarity 21.3%; Pred. No. 1.1;
Matches 178; Conservative 124; Mismatches 272; Indels 263; Gaps 42;

QY 11 GPEVSSNOSNMPII-----NGOI-----ASNSEKSTKASEASPSASSVS 54
DB 501 GPELT-----PYVAALRIYPRYNSGVPEHRTSGNATNAAPTIAVTPPTTLNV 552
QY 55 SWSFSSAKN--ALISLRPAIINKSSPT--DLSQLEASTSTSTYTRVA 101
DB 553 QAISDPITKSMVRVNDLEDDARQGTGPQADIVSYKNGQVVAIDPVEGEYTVTLA 612
QY 102 KDYEAKSNFTPAKSGLENAKTAEYETKMAIDLMALOMERLANS--DPSNHTEEVNN 159
DB 613 RDSQKLSN-NIVQ--VKVAFPLPREAK-----NAVNAAKAKNTALDNNNNLAEKA 664
QY 160 IKKA-LEAKDT-----IDKLNKLVTLQN--QNKSLTEVLKTTSDADQIPAINSOLEINKN 212
DB 665 AKKAAVEAKKAKNTLAGIDDA--KTAAARNAAQKGTTDIAVNPVPYAKPAANAAL--QA 721
QY 213 SADQIITKDERONISEAVLTVNAGEYIKASSEAGIKLQAL--QSIYVDGDSQAVALDA 270
DB 722 AVNKINEISQRPDLTREETKQAFMDQVTRARDAAMAKAVASANNQVTSARDGLNAV--- 778
QY 271 QONNSP-----DNIAATKE-----LIDAEETKVN-ELK 297
DB 779 --NNLETPRAKYPREALGHVQAADAKRQAIRDNANLTAEGQDALRQVDAQTAAEAALN 836
QY 298 QHTGLTDSPLYKKAEEQISOAKDIOETKPSGSDIPYVPSGSAASAGSAAALKSSNN 357
DB 837 QHTTNAT--LAKAD--SDGVKAINDINPQ--PRSKPAQALAEQVAAAKRQAINN 885
QY 358 SGRISLILDDVNNEMA-AI-----ALQGRSMIEQFNVPNPAF----- 394
DB 886 NNQLT-----DEKRAQAIQOYDQALANAKTOYQAANDNNGNQAKTACTTAIINNINPOG 939
QY 395 ---AKELQAMEAQULTAMSQOLVG-----ADGELPAETIATKDALAOLKOPSAD 440
DB 940 TQKAQAIATIEAAEQAKRELQGRNDLTTEERNNLADLTAAQAQAKQAVNOARNTTGA 999
QY 441 GL-----ATMGQVAFPA 453
DB 1000 GAKDNGVAQIQINPTAVYKPDARNAIDQARDKEAFQANTKLDEEKAAMAIKKVQDAA 1059
QY 454 --AKVGGSAGTAGTYQVMVVKQLYTAFSTSSSTAALISQY--SAYTILNLSYSESS 510
DB 1060 RDAKKAIDRAGSNGDVNNAVNO--GKAALQAIK-----ALDSQPSAKTAAAIQONAD 1112
QY 511 GVQASITAPANALSRYSRSGI---ESQGRSA---DASORAAET-IYRDS--QTLGV 560
DB 1113 AKKAAI--TANNAALTOEEKAAMAIKQVEDEAQAQAQAAVDAISRKADYDRAKKQDGLKISY 1170
QY 561 -----YSRLQVLDLSMS--TIVSN---POANOEIMQKL----- 589

1171 PAVQPKLNAIAVDOAATDKKAVINNDTTLTOEKEKAIRKRVDEEAKARQAINDATSN 1230
QY 590 -----TASISKAPOFGYPAYONSADSLQKRAALEREFPVGEESLAEQENA 636
DB 1231 ADVAAKQAGQTOAIVNVPQ--TPAAKNAAKAVEQALADAKKQAIENDPNLTROEKDA 1285

RESULT 37
Q98M03 PRELIMINARY; PRT: 2124 AA.

AC Q98M03.
DT 01-OCT-2001 (Tremblrel. 18, Created)
DR 01-OCT-2001 (Tremblrel. 18, last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, last annotation update)
DE Kinesin-like protein.
GN MR0796.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shlimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002995; BAB48310.1; -
KW Complete proteome.

QY 5 HHMSEGPESVSSNOSNMPIING-----QIASNSET----- 36
DB 261 HSEVNOIERSYSENETRIRSLVDGLGSEAEVYTHAERYRASIAAGAEHLRDEIGAASDI 320
QY 37 -----KESTKAS-----EASPSASSVS-----WSFLSAKNALISLR 70
DB 321 TDSILNASTKLSMTTNSGDTLIDRINESKMSIFDSVGRDITTDKLISTGGEAFASLL 380
QY 71 DALINKNSPTD-----SLQLEASTSTVTRV--AAQDYDEAKSNFDTAKSGLENAKTL 124
DB 381 DTRIAKLFTDTGTLRSLTDLDDRTCKVSLGGAARTLN--SEFEASLNGIE--RTL 435
QY 125 AE-----YETKMAIDLMAALQDMERLANSNPSNNHTEEVANNIKKALEAKD 169
DB 436 AERGQALISEFQTRAEALDTGTOKLNAALEARARQIN---ETLVERAREIAHTFAESKD 491
QY 170 TIDKLNKLVTLQONKSLTEVLKTTSDADQIPAINSOLE-----INKNSA 214
DB 492 T-----LAAMIDQCK--TQT--GADMDIVTSSMLEFARSDFRAGKEAARHYVSSF 541
QY 215 DOIIRDLERONISYEAULTNAGEVIRKASSEAGIKLQALQSIYVDGDSQAVALQAOQNN 274
DB 542 DSDIQRLADARVGEIEAVENHSRKLSSESD-----RMAAAMQADL-- 581
QY 275 SPDNTIAATKELIDAAETKYNELKQHTGLTDSPLYKKAEEQISOAKDIOETKPSGSDIP 334
DB 582 --EKFAESRDGIDAAYT--NQOYKLAEGRS--LIARALEE-----DIRKYNESRAID 628
QY 335 IYVPSGSAASAGSAGALKSSNNSGRISLILDDVDNEMAALQGRFRSMIEQF--NYNN 391
DB 629 -----ASLSHLERIEGRN--RLSLAL-----NEDSGKLVOA--RTIIDEMVAGHWK 673

```

OY 392 PATAKEL--QAMEMOLITAMSQVLGAGGELPAETIAQKD---ALAAOLKRPASDGLATAM 446
Db 674 LAEGRNILSRLEADLGLKLSDRASIDGLVAGOVEKTAEBRAVLAABLE-----722
OY 447 GOVAFAPAAKVGGSAGTAGTVQNMVKOLKYTAFFSTSSSSAAALSGVSAVYKTLNLS 506
Db 723 -----NDIAGISLIEVSHAKLVE-----DRQSLGRSLBEDLSGIRGLDDHS 765
OY 507 ESRGVOYASITOTANPALSR-SYVSRSGLE-----SOGSADASQAAET-----549
Db 766 VKLANDRSLLSOTLEADLAKTAESRSSSIDGLVAGOVEKLEGR--DILKALRESLSTIK 823
OY 550 --IYRDSOTLIDVYSRL-QVLDLSMTSTIVSNPQANOEIMAKLTASIKAPQEGYAYON 606
Db 824 GVYSQSERLEADRGQSLRYEADLOQNVNYSIADHMNRILVQD--RSLSKLLEDDLAKLAD 882
OY 607 SADSLOKRAADLEREPVDC-----ERSLAESQENAFKROPAPITQOVLVYNLSLFS 656
Db 883 SRSSTIDGLVAGOVEKLEAGRIDLRAL-EDALUNTIKTVADSOGLVDDDRAGQA 935

```

ID	Q49526	PRELIMINARY:	PRT:	716 AA.
AC	Q49526:			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1998	(TREMBLrel. 08, Last annotation update)		
DE	lmp1.			
GN	lmp1.			
OS	Mycoplasma hominis.			
OC	Bacteria; Firmicutes; Bacillus/lostridium group; Mollicutes;			
OC	Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID:2098;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MH81;			
RC	MEDLINE-95369882; Pubmed-7543881;			
RA	Jensen L.T., Ladefoged S., Birkeland S., Christiansen G.,			
RT	"Selection of Mycoplasma hominis PC21 deletion mutants by cultivation			
RL	Infect. Immun. 63:3336-3347(1995).			
SM	EMBL; U21963; AAA81014.1;			
SO	SEQUENCE 716 AA; 79864 MW; 28D08C3BC91CB47 CRC64;			

Query Match	6.7%	Score 214	DB 2	Length 716
Best Local Similarity	20.9%	Pred. NO. 0.36		
Best Match 128	Conservative 122	Mismatches 241	Indels 120	Gaps 24

```

0Y 61 SANANLISLROA-----ILKNSSPPDLSLSQLEASSTYSTVTVRYAAKYDDEAKSNEDTA 114
Db 162 SLESTLEIIOATNNKLILKESTREKRONIDNLLAK-----EOLKASISQA 207
0Y 115 KSGLENAKTLAEYETKMAADLAAALQDMERLANSPPSNHTEEVYNNIKALEAQDITDKL 174
Db 208 NOLPQ---LESDNSEIAKAKKASLSDAETKANNAQVASNNTASMQSAKSSIDAAVAELTK- 263
0Y 175 NKLVTLQONNSLSTEVLKTTDTSADQIPALINSQLEINNSAD--QIIKDERONISEAYL 232
Db 264 -KLEFFPNKDRAKENELKQFTN-----QIQEFITNNKNNNTSYLSLSQLTSKRDSKNSVT 317
0Y 233 --TNAGEVIKASSEBGIKLQALQISYDAGDQSGAAVLQAOONNSPNT--AATKELIDAA 289
Db 318 DSNKSDIESANTELQALAKANADKVOADBLAKS--IKEDLNNVSANATLTSKLTLD-- 373
0Y 290 ETKVNELKQEHITGLTDSLPLKKAEEQISOAKQIQEILTPGSGDIPYIPGSGSAASAGSAA 349
Db 374 --KDWTLQOARTEL-----EKVEYAKQAOAIK-----SNNTASMQSAK 408
0Y 350 GALKSSNNSGRISILLDDVDNEMAMAILQGRSMIEOF---NVNNPATAELQAMEQOLN 406
Db 409 SSLDKVAEITKKELETFKKDCAEFNELKQYRRNIOIEITNNKPN---NYSLELSQULT 464

```

QY	407	AMSQJLVG--ABGELPAELQAKKDLAALQPSADJGLATMGQVAFPAALAVGGSGACTAG	465
Db	465	SKROSKNSVTSDSNKSDIESANTELKQALAKANADKQVA--DNLASIKRQLNNVSNN	522
QY	466	TVQNVN-----KOLYK-----TAFSTSSSYAALSDGYSAYKTU	501
Db	523	TL\$AKLTFQDNTIQQAKTELEKEVQKANOQAKSNNTASMQ\$AKSSLD\$AKVAEITKLETF	582
QY	502	N-----SLXSESRGVQ\$A\$ISQTA-NP\$ALSRKSV\$SGIESQGR\$AD\$Q\$RAEITTVRD	553
Db	563	NKDK\$AKE\$F\$NELKQTRN\$QOEIT\$NNKNNPNYS\$ELISQUL-----TSKROSKNSVTD	632
QY	554	SQTLGDVVS-RLOWLDSL\$STIV\$NP\$QAO--E\$EIMOKL\$TASIK\$KAPQ\$E\$P\$AVQ\$NSADS	610
Db	633	SSNKSDIESANTELKQALAKANADKQVADN\$NLASIKRQLNNVSNNATLS-A\$LTFQDNT	691
QY	611	LQKFAAQLERE	621
Db	692	IQQAKTELEKE	702

RESULT 39			
ID	ORG664	PRELIMINARY:	PRT: 1072 AA.
AC	09CF64:		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Unknown protein.		
GN	YOFG OR IL1617.		
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OC	Streptococcaceae; Lactococcus.		
OX	NCBI_TaxID=1360;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=IL1403;		
RX	MEDLINE=21235186; PubMed=11337471;		
RA	Biolotin A., Mincker P., Mauger S., Jallion O., Malarne K.,		
RA	Weissenbach J., Ehrlich S.D., Sorokin A.;		
RT	"The complete genome sequence of the lactic acid bacterium Lactococcus		
RT	lactis ssp. lactis IL1403.";		
RL	Genome Res. 11:731-753(2001).		
DR	EMBL: AE006392; AAK05715.1;"		
DR	PRINTS: PR01608; BACINVASINC.		
KM	Complete proteome.		
SO	SEQUENCE 1072 AA; 113056 MW; 4644462656CA08 CRC64;		

Query Match	6.7%	Score 214	DB 16	Length 1072
Best Local Similarity	19.7%	Pred. No. 0.55		
Matches 154	Conservative 137	Mismatches 280	Indels 212	Gaps 31

[illegible]

```
DB 566 FSNVAEVANNSLASVNNSSSVLSSTSTADNLE--INQGSNLTUKRDSSEISTSGAFISS 623
QY 313 EEOISOAKND-IOEIKPSSGSDPIPIVPGSGAASAGSAGALKSSNSG-RISLLDDVDN 370
DB 624 NOTSEBASSNSMSINSPLSLSTLNSNSGATNQSNSPATYVDNNSSTHSSNIIINSGN 683
QY 371 EMAIALLOGFRSMIEQFNVNNPATAKEIQAMEAQLTAMSDQLVGADGELPAEIQAIKDAL 430
DB 684 DSSD---SDSDSDSDSSNLSSSPNLETNQTISSK-----PSEVNNIISENP 725
QY 431 AALALOPASDGLATAMGQAFAAVAGGSGAGTACTYQM-NYKQLYKT-----AFSS 481
DB 726 KVVSSSNVQENSTHEMSTNPKSSISTSTSSQKESQSNLINTTEGINNPTFFNN 785
QY 482 TSSSSVAALSDGYSAAYKLNS---LY--SESRGVOAISOATANPALRSRVSREGIS 535
DB 786 SSEBSASASILTYSNNNSSESETEGCLYISNQAQNDNGSEIHSLSPPSSNSENNSIOS 845
QY 536 Q-----GRSADASQRAAETIV-----RDSQT-----LGDVYS----- 562
DB 846 QALLESSKSTNRSSSLIISTHSPQMEDNQNSDEVKNNNVEISILQNLNISTNKT 905
QY 563 -----RLOVLDLSMS-----TIVSNPQAND-EIMOKLTA---S 592
DB 906 HNSLTSQKLSYITLPSKSKVTNEKNENSNTVSEKILKTQKNDSCNLQITALDLS 965
QY 593 ISKAPQFGYPAYQNSDSLQKFAOLEREFEVDGERS-----LAESQENAFRKQAPQIO 645
DB 966 FKK-----EYETMEDSSTVDPKVLNDN--NGDRSQNNKTJTIKKDKKVFYFKRSEFNS 1016
QY 646 QVL 648
DB 1017 KIL 1019

RESULT 40
Q8T805 PRELIMINARY; PRT: 1489 AA.
AC Q8T805;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE SD05887P.
GN CG3493.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuncio J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celiker S.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY075577; AAL68382.1;
SQ SEQUENCE 1489 AA; 170145 MW; 876DC2DE469945AF CRC64;
```

Query Match 6.6%; Score 213; DB 5; Length 1489;
Best Local Similarity 19.1%; Pred. No. 0.98;
Matches 141; Conservative 127; Mismatches 288; Indels 184; Gaps 23;

```
QY 5 HHMHSGPESVSNQSNMPTINGQIASNETKESTKASEASPSASSVSWSFSSAKN 64
DB 326 HTELEKQTEVFKLQEKLEKQLESQRESHNNEVKEQFKLQATKQEVDAK-----LMATEH 380
QY 65 ALISLDAIILNNSSPTDLSOLEASTSTSTYTRV--AAKDYDEAKSNFDTAKSGLENAK 122
DB 381 ILNLTKESTVAKKEQGVLTLEQLA-----IRVENEQKVKDLQKQNDNRNTQASDSE 433
```

```
QY 123 TLAEYETKNADIMAILQDMERLANS----- 147
DB 434 QLKTLQAAVOADESOLLKSKDQLLESIRSEQAQKQOLKHLKQOLKQKQENENYLDKLE 493
QY 148 --DPSNNHTEEVNNIKKALEAKQKDTDKLNKY-----TIQONKS-----LTEVL 191
DB 494 SKRSSDQTENEAODQOKTLQAAKDEAE--SKLLATEELLHSIRNDYKQAEKVALLDEKL 551
QY 192 KTTD-----SADQIPALNSQLEINKNASADQITKDL-----ERONISYEAVLTNAGEYI 239
DB 552 KTLSEKNVYVNEKLIHINQREAOQSTDSQKTNELRAKDEAFALISTEHSILNLQAL 611
QY 240 KASSEAGIKLQAL-----QSIYDA--GDQSOAAVLAQAOQNNSPDNIAATELIDA 288
DB 612 SAKERQASLEQSLNALKTESEHSIQDLRLHNDQILEIYQRRHQQDWEAQLARAREELAA 671
QY 289 AET-----KNELKQEHGTGLDSPLYKKAE--QISOAKQDIQKIPSSDPI 335
DB 672 IQSQRHLALELEKSLMEERESVAALNSEKASQEOHRUKLEQILQREIQILODQHAN-- 728
QY 336 VQPSGSAASAGSAGALKSSNNGRISLLDDVDNEMAAIALQGRSMTEQFPVNNPATA 395
DB 729 -SESETVAAIKQLEALSDQLATSQASILLAKEKEKELKASGNKLNKIKKHQHQAKSSDS 787
QY 396 KELQAMEAQLT--AMSDQLVGADGELPAEIQAIKDALAQALKQPSADGLATAMGQVAF 452
DB 788 VLEALQSLADRLSHSRQVESEKEELQARVGLIEI----- 825
QY 453 AAKVGGSGTGTQOMNKKLYKTAFSTSSSSVAALSDGYSAVKTLSLXSE----- 507
DB 826 -----GTWQAOQOV-----ODSHSELERERKKLESRESLQOQVDS 864
QY 508 ---SRGVOAISQATANPALSR-----SVSRSGIESQGRSADASORAAETIVRDSQTLGD 559
DB 865 AADERTSAKLEIEIOSENTYLAERNCLLEBQANHLESQ---LOAKQDEIGIAKIQOVLDD 921
QY 560 VYSRLQVLDLSMS---TIVSNPQA--NOEIMOKLTAISKAPQFGYPAYQNSDSLQK 614
DB 922 EHSKIQNAQELMDHHRITLQDKDAYEKDKLTKHTLD-----CLQASASELHRY 971
QY 615 AAOLEREFPDGERSLAESQ 634
DB 972 KANLDREIKEDQDQOLSEURE 991
```

Search completed: March 12, 2003, 12:21:40
Job time : 133 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 09:42:47 ; Search time 21 seconds
(without alignments)
1303.542 Million cell updates/sec

Title: US-10-007-693-139
Perfect score: 3204
Sequence: 1 MHNNHMHESGPESVSNOS.....PAFIQOVLVNIASLPSGYLS 660

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224.5	7.0	1790	US01_YEAST	P23586 saccharomyc
2	209	6.5	1433	REST_CHICK	O42184 gallus gall
3	208	6.5	1306	MSB2_YEAST	P32334 saccharomyc
4	202.5	6.3	1966	MYSB_CAEEL	P02566 caenorhabdi
5	201	6.3	2116	MYSD_DICDI	P08799 dictyosteli
6	200	6.2	1969	MYSA_CAEEL	P12844 caenorhabdi
7	199.5	6.2	1140	YMB6_YEAST	Q04893 saccharomyc
8	199.5	6.2	1381	YBE7_YEAST	P34216 saccharomyc
9	199.5	6.2	1957	YDB6_SCHPO	Q10411 schizosacch
10	199	6.2	1509	MYSN_ACACA	P05659 acanthamoeb
11	199	6.2	1940	MYH3_RAT	P12847 rattus norv
12	199	6.2	2492	TALA_DICDI	P54633 dictyosteli
13	198.5	6.2	539	MYH3_YEAST	P39922 hydra atten
14	197	6.1	1938	MYS_AFOIR	P24733 aequipecten
15	194	6.1	1093	TMF1_HUMAN	P82094 homo sapien
16	194	6.1	1940	MYH3_HUMAN	P11055 homo sapien
17	194	6.1	1962	MYSA_DROME	P05661 drosophila
18	193	6.0	1938	MYHD_HUMAN	O94kx3 homo sapien
19	192.5	6.0	1727	ALML_SCHPO	O94kx3 schizosacch
20	192	6.0	1938	MYH4_RABIT	Q28641 oryctolagus
21	192	6.0	2022	ANT1_ONCVO	P21249 onchocerca
22	190.5	5.9	1462	NKCR_HUMAN	P30414 homo sapien
23	190	5.9	1939	MYH1_HUMAN	P12882 homo sapien
24	189.5	5.9	1938	MYSD_CAEEL	P02567 caenorhabdi
25	188.5	5.9	3210	CEMF_HUMAN	P49454 homo sapien
26	188.5	5.9	918	YMOB_CAEEL	P34487 caenorhabdi
27	188.5	5.9	995	YIO9_YEAST	P40442 saccharomyc
28	188.5	5.9	2017	MYSN_DROME	O99323 drosophila
29	188	5.9	1528	SPAA_STRMO	P19179 streptococc
30	187.5	5.9	1935	MYSS_CYPGA	O90339 cyprinus ca
31	187	5.8	1084	MYSS_RABIT	P02562 oryctolagus
32	187	5.8	1935	MYH7_PIG	P79293 sus scrofa
33	186	5.8	697	MEP1_LYCCE	P93203 lycopersico

ALIGNMENTS

RESULT 1	ID	US01_YEAST	STANDARD:	PRT:	1790 AA.
AC	P23586;				
DT	01-MAY-1992 (Rel. 22, Created)				
DT	01-MAY-1992 (Rel. 22, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Intracellular protein transport protein US01.				
GN	US01 OR INT1 OR YDL058W.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=X2180-1A;				
RX	MEDLINE=91185402; PubMed=2010462;				
RA	Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,				
RA	Yamasaki M.;				
RT	"A cytoskeleton-related gene, us01, is required for intracellular				
RT	protein transport in Saccharomyces cerevisiae.";				
RL	J. Cell Biol. 113:245-260(1991).				
RN	[2]				
RP	SEQUENCE OF 782-1790 FROM N.A.				
RA	Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,				
RA	Kendrick K.E.;				
RL	Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 1-8 FROM N.A.				
RA	Bai Y., Symington L.S.;				
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.				
CC	- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI				
CC	COMPLEX.				
CC	- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR				
CC	MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE				
CC	ER AND THE GOLGI COMPLEX.				
CC	- DOMAIN: THE RODLIKE TAIL. SEQUENCE IS HIGHLY REPETITIVE, COMPOSED				
CC	OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL				
CC	COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.				
CC	- SIMILARITY: BELONGS TO THE VDP/US01/YBL047C FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: X54378; CA38253.1; -				P35749 homo sapien
DR	EMBL: L03188; AAB00143.1; -				Q28628 oryctolagus
DR	EMBL: U53668; AAB66559.1; -				Q10970 mycobacteri
DR	PIR: A38455; A38455.				P30622 homo sapien
DR	SGD: S0002216; US01.				P12845 caenorhabdi
DR	InterPro: IPR002017; Spectrin.				P47033 saccharomyc
DR	Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.				O94623 homo sapien
KW					P35580 homo sapien
					O92351 schizosacch
					P23504 streptococc
					P02564 rattus norv
					O27991 bos taurus

```

FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

```

Query Match Best Local Similarity 20.9%; Score 224.5; DB 1; Length 1790; Matches 155; Conservative 130; Mismatches 268; Indels 189; Gaps 32;

```

QY 3 HHHHHHSGPESSY-----SSNOSMN-----PIINGIASNSE--- 35
DB 969 NNTKMDQAEKESLITKAVEESKNESSIQLSNLQKIDMSQEKENFOIERGSIKTEIQLK 1028
QY 36 ---TKESTKASPEASPSASSVSSWFLSSAKNAL---ISLRDAIINKSSPDTLSOLE 88
DB 1029 KTIISDLEQKREIISKSDSKDEYESQILKKELETAATANDENVKISLTKTREBLE 1088
QY 89 ASIST-----SYTPRAAKYIDAKSNFTPAK--SGLENATLAIEFKMADLMAALO 139
DB 1089 AELAAAKNKNLELETKLETSKALKKEVEHEHLKEEKIQLEKAEYEQOOLSLFANLE 1148
QY 140 DMRLANSPDSNHTPEVNNIKALE--AOK-----DTIDKLN-KLVTLQNKSLTEVL 191
DB 1149 SLEK-----EHEDLAQLKKEEIOIANKEQYNEISQLDETTSTQOEMESIKR-- 1198
QY 192 KTTSDAQDIPAINSS---OLEINKNSADQI---IKDLERONISYEAVLTNAGEVIRASSE 244
DB 1199 KNDLEGEVAKMSTSEEGSNLKKESEIDALNQLKELKKNETNEASLT---ESIKSVES 1255
QY 245 AGRIKIQ-----ALQSYDAGQSQAAVLAQANNQNSPDINATKELIDAAET 291
DB 1256 EYVAKIKELDECNFKKEVSELEDKLKASDKSKYLEIQKESSE---KIKELELAKTT 1310
QY 292 KVNELKOEHTGL-----TDSPLVKKAKEQISOAKQIDQIEKSGSDIPI 335
DB 1311 ELKIQLEKINTNLKAKEKSESELRLKTTSSBEKKAHEQLKKNELQ--IKNQAE--- 1366
QY 336 VGPSSGASASGAAGALKSSNNSGRISLLIDV---DNEMAAIALQGFMSIEQFNVN 390
DB 1367 --KERKLLNNGSST---ITQYSEKINTTLEDELIRLQNNENELKAKEIDNRSELEKVSLS 1421
QY 391 NPA-----TAKEIQ--AMEAQITAMSDQVAGDELPAITQAKDLAALQPSA 439
DB 1422 NDELEEKQNTISLQDELISYKDKITRNDKELTIERDKRKRLESEKQLRAQE--- 1477
QY 440 DGLATAMQVAPAAKAGVAGSGAGTAVQANNVQKYTAASSTSSSYAALISDGYSAKK 499
DB 1478 -----SKAYVEE-----LKKL-----EESSEKREKLEKSEMMK 1508
QY 500 TLNLSYSESRSGVQASISQITANPALSRVSRSGIESQGRSADASQRAETIIVDSQ--TLG 558
DB 1509 KLSSTIESNETELKS-----SMETIRKSDEKLEQSKKSABEDIKNLQHEKS 1554
QY 559 DVSURL-----QVLDLSMS--TIVSNPOANOBEIMQKITASISAPQFGYAVONSADSLQ 612
DB 1555 DLISRISEKDEIEELKSKLRIDAKSGSELETVQKOLNNAQEK---IRINAENTV--LK 1609
QY 613 KFAAOLEREFVDEGRSLAESQ 634
DB 1610 SKLEDIERELKQKQAEIKSNQ 1631

```

```

RESULT 2
ID REST_CHICK STANDARD: PRT: 1433 AA.
AC 042184: 042228: 057563: 057564:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Restin (Cytoplasmic linker protein-170) (CLIP-170).
GN RSN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=90311;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98137792; PubMed=9469933;
RA Gripapic L., Volosky J.M., Keller T.C. III;
RT "Cloning and expression of chicken CLIP-170 and restin isoforms.";
RL Gene 206:195-208(1998).
RN [2]
RP SEQUENCE OF 17-1139 FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=pectoralis muscle;
RA Gripapic L., Keller T.C. III;
RT "Identification and expression of two novel CLIP-170/Restin isoforms expressed predominantly in muscle.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE CYTOSKELETON (BY SIMILARITY).
CC - ALTERNATIVE PRODUCTS: At least 4 isoforms; 1 (shown here), 2, 3/CLIP-170(11) and 4/CLIP-170(11+35); are produced by alternative splicing.
CC - SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC
DR EMBL: AF014012; AAC60344.1; -
DR EMBL: AF020764; AAC60345.1; -
DR EMBL: AF045650; AAC03547.1; -
DR EMBL: AF045651; AAC03548.1; -
DR InterPro: IPR000938; CAP-GLY.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam: PF01302; CAP_GLY; 2.
DR SMART: SM00343; ZnF_C2HC; 1.
DR PROSITE: PS00845; CAP_GLY_1; 2.
DR PROSITE: PS0245; CAP_GLY_2; 2.
KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
FT DOMAIN 79 121 CAP-GLY 1.
FT DOMAIN 144 207 SER-RICH.
FT DOMAIN 235 277 CAP-RICH.
FT DOMAIN 305 332 SER-RICH.
FT DOMAIN 351 1353 COILED COIL (POTENTIAL).
FT DOMAIN 1414 1427 CCHC-BOX.
FT VARSPLIC 458 492 MISSING (IN ISOFORM 2).
FT VARSPLIC 458 492 TOTKLEHAIKIKELQSLFEKTKADKLORELEDTR ->
FT VARSPLIC 458 492 S -> GGSSKVS (IN ISOFORM 3).
FT VARSPLIC 803 803 T -> RKRISDEPENT (IN ISOFORM 4).
FT VARSPLIC 458 458 K -> R (IN REF. 2; AAC03547).
FT CONFLICT 309 309 E -> V (IN REF. 2; AAC03548).
FT CONFLICT 440 440
SQ SEQUENCE 1433 AA; 161026 MW; 5631CE8683498E23 CRC64;

```

Query Match Best Local Similarity 17.5%; Score 209; DB 1; Length 1433; Matches 154; Conservative 126; Mismatches 260; Indels 342; Gaps 28;


```

Db 302 MSAAGQIASSSSTDNFTMSEFTSLSTEVDSVSTVALISAPFLQ--TSTNSFSIV 359
OY 391 NPATAKELOAMEAQLTAMSDQLVAGDELPAEIOAIKDALQAKOPSDGLATANGOVA 450
Db 360 SPSVSVSPSSSDVASSSTANV-----VSSFSDFPPTSTSGSVVSAQSA 407
OY 451 FFAA-----KVGGSAGTAGVQMNKYKTAFFSSTSSSYAALSD---GYSAKYTL 501
Db 408 SALAFQSSSTEVYGCASASTMSSTLSLTSTLSQSTPLDSSSLASSASSDLDYGVSTASI 467
OY 502 NGLYSRSRGVQSAISQTNPMALSRSVSRGIESQGRSADAGRAAETVRSQTLGDVY 561
Db 468 PLLSASEQSTSSSFS-VVSPSVSEVPSQS-----SSDVASTAPSVSSSFS---Y 515
OY 562 SRLQVLDSTLM-----STIVSNPOANEELMOKLTASISKAPQFGYPAYONSADSLQFAA 616
Db 516 TSLQAGSSMTNPSSTSTIYSSSTGSSSESAASTASATL-----SGSSSTYMAAG 564
OY 617 QLERFVDEGRSLAESQ 633
Db 565 NLSQSPSTSLISESQ 581

RESULT 4
MSB_CAEEL
ID MYSB_CAEEL STANDARD: PRT; 1966 AA.
AC P02566;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain B (MHC B).
GN UNC-54 OR MYO-4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83273600; PubMed=6576334;
RA Karn J., Brenner S., Barnett L.;
RT "Protein structural domains in the Caenorhabditis elegans unc-54
RT myosin heavy chain gene are not separated by introns."
RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
RN [2]
RP SEQUENCE OF 850-1966 FROM N.A.
RX MEDLINE=82272395; PubMed=7202124;
RA McEachlan A.D., Karn J.;
RT "Periodic charge distributions in the myosin rod amino acid sequence
RT match cross-bridge spacings in muscle."
RL Nature 299:226-231(1982).
RN [3]
RP SEQUENCE OF 1876-1966 FROM N.A.
RX MEDLINE=8332892; PubMed=6571695;
RA Mills N., Gesteland R.F., Karn J., Barnett L., Bolten S.,
RA Waterston R.H.;
RT "The genes sup-7 x and sup-5 III of C. elegans suppress amber
RT nonsense mutations via altered transfer RNA."
RL Cell 33:575-583(1983).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MYOSIN MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN

```

```

CC C.ELEGANS.
CC -1- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY
CC WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: J01050; AAA28124.1;
DR EMBL: V01494; CAA24738.1;
DR PIR: A02992; MWMN.
DR HSP: P08799; 1MND.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00242; MYSC; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 850
FT DOMAIN 851 1966 COILED COIL (POTENTIAL).
FT DOMAIN 851 1164 ALPHA-HELICAL TAILPIECE (S2).
FT DOMAIN 1165 1176 HINGE.
FT DOMAIN 1165 1966 LIGHT MEROMYOSIN (LM).
FT NP_BIND 177 184 ATP (BY SIMILARITY).
FT DOMAIN 665 687 ACTIN-BINDING.
FT DOMAIN 769 783 ACTIN-BINDING.
FT MOD_RES 128 128 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 705 705 ALKYLATION (SH-1).
FT MOD_RES 715 715 ALKYLATION (SH-2).
FT CONFLICT 1337 1337 E -> L (IN REF. 2).
FT CONFLICT 1880 1880 I -> L (IN REF. 2).
SQ SEQUENCE 1966 AA; 225125 MW; B6f0BB2FE27B67F CRC64;

Query Match
Best Local Similarity 21.9%; Pred. No. 0.33;
Matches 136; Conservative 112; Mismatches 244; Indels 129; Gaps 27;

OY 82 DLSQLEASTSTSTYRVAKDYDEAKSNFDTAK-----SGLENART-LAEYETKADLMA 136
Db 866 DKVKALEDSLAKEEKLK--KELEESSAKLVEEKTSLFTNLESTKTQLSDAEERLAKTEA 922
OY 137 ALQDM-----ERLANSPPSNHTEEVNNIKKALEAQKDTIDKLNKLVTLQONKRSUT 188
Db 923 QOKDASKQSELNDQLADNE---DRTADVQRAKKTEAEVLEALK-----QODELSMR 974
OY 189 EYLKTTDSAD-QIPAINSOLEINKNSADQITKDLERONISYEAVLTNAGEVITAKASSEAG- 246
Db 975 KAESEKQSDHOIRSLQDEDMQ-----QODEAIKLINKEKHQHEINRKLMEDEQSEEDKGN 1030
OY 247 -----IKIGQALQSLVADGDQSQ--AAVLQAQONNSPNIATKELIDIAETK-----VN 294
Db 1031 HONKVKAKLEQTLDDLESLERKRRARADLDQKRRVEGELTIAEDNIDESQRORHLDEN 1090
OY 295 ELKQEHTEG-----TSPVKKRKEPQISOAKDIOEIKRSGSDIPVGRSSAASAG 346
Db 1091 NLKKKSELSHVSRLDEQALYSKLQROIKQGOQKISLEEB-----LENERQSRKAD 1145
OY 347 SAAGALKSSNNGRISLDDVDNEMAAITALQ-----GFRSMIEQFNVNP-- 392
Db 1146 RAKSDQLQ-----ELEELGEKIDEGGATAAQVEVKKREAEELAKRLRLEEFANNNHMQ 1200
OY 393 --ATAKELOAMEAQLTAMSDQLVAGDELPAE-IOAIKKA--LAQALKOPSDGLATANG 447

```

```

Db 1201 LGLKKHTDAVALTDLQDLNKKAKYKDKAQAARDPAEDLAAQLDQETSGKL----- 1255
OY 448 QVAFAAKVGSGAGTACTGVQNNKOLYKTAFSSSTSSSSAAALSDGYSAYTLNLSYSE 507
Db 1256 -----NNEKLAKOFELQLTEL-----QSKADQSQQLQDFTSLKRLHSENGD 1298
OY 508 SRSGVQASISQTPALNSRSHSGIESGGRSADASORAAETIVDSQTLGDVYSRLQYL 567
Db 1299 LVROLEDAESGVNQLTRLSQLSQTLSEARTRADEARQRQVVAQAQKNYQHAEQLQ-- 1356
OY 568 DSLMTIYSNPQANDEITMOKITASIKAPQGFYPAVQNSADSLQKFAAQLEREFVQGER 627
Db 1357 ESLEEEI-----EGKNELRLQL-----SKA-----NAD-IQWKARFE-----GEG 1391
OY 628 SL-AESQENAFKQAPFIQOV 647
Db 1392 LKKADELDAKRRQAKINEL 1412

RESULT 5
MYS2_DICDI STANDARD: PRT: 2116 AA.
AC P08799;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycelozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092266; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Leitwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from
   Dictyostelium discoideum.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN=AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lueck-Vietmeier D., Schleicher M., Grabatin B., Wippler J.,
   Gerlsch G.;
RT "Replacement of threonine residues by serine and alanine in a
   phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerlsch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
   Dictyostelium myosin heavy chain.";
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Suton K., Holden H.M.,
   Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
   discoideum complexed with MgADP, BeFx and MgADP.Alp4-.";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
   truncated head of Dictyostelium discoideum myosin to 2.7-A
   resolution.";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;

```

```

RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
   Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammaS, and MgAMP-PNP complexes
   of the Dictyostelium discoideum myosin motor domain.";
RL Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bageshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization
   of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
   Dictyostelium discoideum myosin motor domain.";
RL J. Mol. Biol. 274:394-407(1997).
CC -I- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
   ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -I- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
   INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHCI), 2 ALKALI
   LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
   (MLC-2).
CC -I- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
   CORTEX.
CC -I- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
   MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
   SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
   SUBFRAGMENT (S2).
CC -I- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
   CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
   CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -I- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
   THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC -I- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
   ACTIVITY. PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
   POSITION (688).
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M14628; AAA3327.1; -.
DR PIR: A26655; A26655.
DR PIR: S00250; S00250.
DR PDB: 1MMA: 03-DEC-97.
DR PDB: 1MMD: 17-AUG-96.
DR PDB: 1MMG: 03-DEC-97.
DR PDB: 1MMN: 03-DEC-97.
DR PDB: 1MND: 17-AUG-96.
DR PDB: 1MNE: 17-AUG-96.
DR PDB: 1VOM: 23-DEC-96.
DR PDB: 1LVK: 28-JAN-98.
DR Dictydb: DD01008; mhca.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR001609; Myosin_head.
DR Pfam: PF00063; myosin_head. 1.
DR Pfam: PF00612; IQ. 2.
DR Pfam: PF02736; Myosin_N. 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head. 1.
DR SMART: SM00015; IQ. 1.
DR SMART: SM00242; MYSC. 1.
DR PROSITE: PS50096; IQ. 1.

```



```

FT DOMAIN 857 1969 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP.
FT DOMAIN 667 689 ACTIN-BINDING.
FT DOMAIN 770 784 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (TR1-) (POTENTIAL).
FT MOD_RES 707 707 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 717 717 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1969 AA: 225509 MW: 64577BBAF7EAD80A CRC64:

Query Match 6.2%; Score 200; DB 1; Length 1969;
Best Local Similarity 18.9%; Pred. No. 0.42;
Matches 142; Conservative 152; Mismatches 304; Indels 154; Gaps 28;

QY 28 GGIASSEKREKTKASEAPSSASSVSSW-SFLSSAKNAL-----ISLRALINKNSPTDS 83
DB 863 GELAVKIQLE--EAVQGEIARISQLESOVADYVEKKNLFSLTEKKNLDAEERNK 920
QY 84 LSOLEASTSTVTRVAADYDEAKSNFDTAKSGLENAKTLAEYETKMDLMAALQDMER 143
DB 921 LNLQKA-TLESKLSDTGTGLEDMQERNEDLARQKKTDDELSDTKKHVQDDELSLRKAO 979
QY 144 -----LANDPS-----NNITEVY-NIKKALEAKQDTIDKLNKL-- 177
DB 980 EKQSRDHNIRSLQDEKANDQEVAVKLNKEKKHQEESNRKLNEDLQSEEDKVNHLERIK 1039
QY 178 -----VTLQONKSLTEVLYK-----TTSDAQDIPA----- 202
DB 1040 LEOQDELEENIDREKRSKGLIEKAKRKVEGDKLYAQENIDETTKOKHDEVTLLKREED 1099
QY 203 ---INSOLEINRN---SADQIIKLERONISYEAVL---TNAGEYIKASSEAGIKQOA 252
DB 1100 LHHTNAKLANNENSIKAKLRLKELTARNALBEELEARNRSKOKSDRSSEARLELEL 1159
QY 253 LOSYDAGQSOAAVYLOAQONNSPDNIATKELIDAA---ETKVNELKOEH---TGLTD 305
DB 1160 TRLRLQOQG-ATPAOLEAKKRKEAEIAKLRLKEEDSLNHETALISLRKHGDSVALTE 1218
QY 306 -----SPLYKKAEQISOQKDIQOE-----IKPSSGSDIPVGS 339
DB 1219 QLETLQKLAKKSEAEKSKLQRLDEESQAHATDESVRSQRLKALKTIEVQYSELQTKADE 1278
QY 340 GSAASGSAAGALKSSNNSGRISSLDDVDNEMAAI--ALQGFMSMIEQFNVN-- 391
DB 1279 GSRQLODFALALRNENNSDLNRSLEMDQNLNHLRKSTLQSLDLTRRYDEESRE 1338
QY 392 ----PATKELQAMEAQLTAMSDQLVGADGELPAETIQAIKDALQAKQPSADGLATANG 447
DB 1339 RQALATATANKLHEHTILREHLDDEAEKADLTLROIISKLNAEIQOKKARFDSGL-NKLE 1397
QY 448 QVAFPAAKVGGSGAGTAGVQNMVKQLYKTAFSSTSSSYAALSDPGYSAKYLNLNLYSE 507
DB 1398 EIE-AAKK-----ALQKLVQEL-----TDINEGIFAKIASQEKVFKIMQDL-DD 1440
QY 508 SRSGVQSALISQANPALPSRSVRSRGIESQGRSADASQRAETIIVRDSQTLG-DVYSRLQV 566
DB 1441 AOSDVEKAAQAQVAFYEKHRQFESIIEWKKTIDLSSELDAQAQRNRLQSLDLFRAKTA 1500
QY 567 LQSLMSTIYSNPQANE--EIMOKLTASISKAPQEGYPAVQNSADSLQFAAQOLEEFPD 624
DB 1501 NQELAEVYLDSTRRKNSLQAEKYLTDQEGEGR-----SVLELQKIVKLEVEKEE 1552
QY 625 GERSLAESQENAFRKQAPFIOOVLVNIALSFS 656
DB 1553 LQKALDEA-EALAEAEKAVLRAQIEVSOIRS 1583

RESULT 7
YM96_YEAST STANDARD: PRT: 1140 AA.
AC 004893:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)

```

```

DE Hypothetical 113.1 kDa protein in PRE5-PET4 intergenic region.
GN YMR317W OR YM9924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_Taxid=49312;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=528bc / AB972.
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.,
RA Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
CC -I- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: 254141; CAA90835.1; -.
DR SGD: S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA: 113070 MW: 0153EBCA24FE5427 CRC64:

Query Match 6.2%; Score 199.5; DB 1; Length 1140;
Best Local Similarity 19.1%; Pred. No. 0.23;
Matches 124; Conservative 109; Mismatches 278; Indels 139; Gaps 17;

QY 12 PSVSSNOSNMPI-----INGQIASNEKREKTKASEAPSSASSVSSMSFSL 60
DB 64 PEFTSSSLSDPTIASILSESLVIFSSLTSTSDISSTVNDVSSSTSGSNYSALS 123
QY 61 SAKNALISLRDALINKNSPTDLSQLEASTS--TSTVTRVAADYDEAKSNF--DPAK 115
DB 124 SP-----NAQLSSSTETDTSISSAQTSPQTSNGGSGSSEPLKRSVLETPAS 175
QY 116 SLENAKTLAEYETKMDLMAALQDMERLANDPNNHTEEVNNIKKALEAKQDTIDKLN 175
DB 176 SSDTTAVVSTPTT-----LTDVSSSPKISSGSAVTSVGTTSDAKRE----- 218
QY 176 KLVTLQONKSKLTEVLYKTTDS-----ADQIPAINSOLEINKNSADQIIKLERONISYE 229
DB 219 ---VFSSSTSDVSSLSLSTSSPASSTISSETLPSSITLSTSPVSSSEAPKATSSVSSE 275
QY 230 AVLTNAGEYIKASSEAGIKRLQALQSYDAGDOQAAYVLOAQONNSPDNIATKELIDAA 289
DB 276 ASSSTSSSV---SSEAPL---ATSSVSSSEAPSSSTSSVSSSEAPSSSTSSVSSEISPT 328
QY 290 ETKV-----NELKQEHITGLDPSLYKKA-----EQQISQA 319
DB 329 SSSVSSSEAPLATSSVSSSEAPSSSTSSVSSELSSTSSVSSSEAPLATSSVSSSEAPST 388
QY 320 QKDIOEIKPS--GSDPIYIPGSAAGSAGALKSSNNSGRISLLDDVDNEMAAIAL 377
DB 389 SSSVSSSEAPSSSTSSVSSSEAPSSSTSSVSSSELSSTKSSVMSSEVSAATSLVSSSEAPAI 448
QY 378 QGFMSMIEQFNVNPNATKELQAMEAQLTAMSDQLVGADGELPAETIQAIKDALQAKQPS 437
DB 449 SSLASS-RLFSSKNTSVSTLVATEASSVTS-----LAP 482
QY 438 SADGLAT-----AMQVAFPAAKVGGSGAGTAGVQNMVKQLYKTAFSSTSSSYVA 489
DB 483 SSETLASNLSISSLSTGTNSTVSTTSAASSTLSKSSVSNRAVSTKSTSSSDLSKS 542
QY 490 ALSDYSAYKTLN-----SLYSESRGVQSALISQANPALPSRSVRSRGIESQGRSADASQ 544
DB 543 SVTFPNSSTVTTSPASISLITASPLPSVWSDLTSSSEASSISNLSASSPDDNNNTIASA 602
QY 545 RAEETIVRDSQTLGVYSHRLQVLDLSMSTIYSNPQANEIIMOKLTASIS 594
DB 603 SLIVTKTKNS-----VVSSIVSSITSTSETTNESNLATSTSLIS 641

```


RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Meltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Mostl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forzburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.",
RL Nature 415:871-880(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 270690; CAA94624.1; -
CC Hypothetical protein.
KW
SQ SEQUENCE 1957 AA: 222785 MW: 3F480CA06171D9DA CRC64;

Query Match 6.2%; Score 199.5; DB 1; Length 1957;
Best Local Similarity 19.0%; Pred. No. 0.44;
Matches 144; Conservative 132; Mismatches 279; Indels 201; Gaps 28;

OY 23 NPITNOIASNSTEKSTKSEASPSASSSVSSMSPSSAKMALISRDAILKNSPTD 82
DB 153 NGILISQULSSNKKDKMT-----SSVTLTSEEDVSYQKRLTNESNFAKQSPAYD 205
OY 83 SLTSLSEASTSTVTR---VAAKDYDEAKSNFDTAKSGLENAKTLAEYETKMDLMAAQ 139
DB 206 LSNQL-----LVTEKLDDKEKEDYEKIKEDVSIKA-----SLAEQASNNKSLKEQGE 253
OY 140 DMRILANSDBSNHTEEVNNIKKALEAKQDTI-DKLNKLVTLQONKSLTEVLEKTTDSAD 198
DB 254 RLEKLIVS--SNKTVSTLRQTFENSLRAECKTLOEKLEKCAINEDEDSLLELKHNV---- 307
OY 199 QIPAINSOLEINKNSADQIIKDLERONISYEAULTNAGEYTKASSENGITLGGALOSIVD 258
DB 308 ---ANVSDAIVHK---DKLIEDLSTR-ISEEDNLIKSERDLSIKNE---KLEKILRLNTIG 357
OY 259 AGDQSAALVLAQOQNN---SPDNIATKRELIDAEFKVNLKQEHGTLTDS----- 306
DB 358 SLKSDSTNSQLEEWELKESNRTHSQULTD-AESKLSFEBQNKSLKSIDYQNNLS 416
OY 307 ---PLYKAEQISQAQ-----KDIOIKIPSGSDIPIVPGSGSA 342
DB 417 SKDKMVKQVSSQLEEARSLAHATGKLAETNSERDFONKRIKDFEKIEQL-----R 468
OY 343 ASGSAAGALXKSSNNGRISLLDDVDNEMAAIALALOGFRSMIOFNWNNPATAKEIQAME 402
DB 469 ACINSSSNELKEK-----SALIDKRDQE-----LNNLRQIKQKRVSESTSSLSQISQ 517
OY 403 AOL-----TAMSDQLVGADGELPAEIQAIKDALAQAOLKQPSADGATMGQVAFAAK 455
DB 518 ROLLNKKKKHVEESQLENEKGELOTEIS-----NSEHLSOLS--TLAAEK 562
OY 456 VGGGAGTACTQVMNVKQLYKTAFFSTSSSSVAAA-----LSDGYSAYTKTINLSYSESR 509
DB 563 EAAVATNNELSESKNSIQTLCAFOEKLAKSVQMLKENEQNFSLDTSFKKLINESHQELE 622
OY 510 SGVQSAISQ---TANPALSSVSRSG-----ISEQGSADAQSR 545
DB 623 NNHOTITTKQDKTSSKIQLOLERANFEQEKSTLSDENNDLRTKLLKLEBSNKSLLIKQOE 682

OY 546 AAEIVRDSOTIGD-----VYSRL-----QVILSL----- 570
DB 683 DVDSLEKNITQTLKEDLRKSEALRFSKLEKKNLEIYDINKGHEFLQORDHJSSLSLD 742
OY 571 ---MSTVSNPQANOBEINOKLTASFI-----SKAPQFGYPAYVNSADSLQKFAQLERE 621
DB 743 AKNTNALISSELTSSSEDEVRYRLTANVETLQDSKAKKQSPSLVNSYQGISNLNHELRRD 802
OY 622 FV-----DGEISLAESQANRKPAPFIQOV 647
DB 803 HVNMQNNNTLSESKSLATDCEINLTQNNMTLIDNV 838
RESULT 10
ID MYSN_ACACA STRAND: PRT: 1509 AA.
AC P05659;
DT 01-NOV-1988 (rel. 09, Created)
DT 01-NOV-1988 (rel. 09, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308395; PubMed=3040773;
RA Hammer J.A. III, Bowers B., Paterson B.M., Korn E.D.;
RT "Complete nucleotide sequence and deduced polypeptide sequence of a
RT nonmuscle myosin heavy chain gene from Acanthamoeba: evidence of a
RT hinge in the rodlike tail.";
RL J. Cell Biol. 105:913-925(1987).
CC -I- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -I- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
CC (MLC-2).
CC -I- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC OF ALPHA-HELICAL STRUCTURES. THIS REGION IS
CC INTERRUPTED BY A HINGE AND JOINED BY A NONHELICAL TAILPIECE WHERE
CC THE REGULATORY PHOSPHORYLATION SITES RESIDE.
CC -I- MISCELLANEOUS: THE HINGE REGION MAY PLAY A KEY ROLE IN MEDIATING
CC THE EFFECT OF HEAVY CHAIN PHOSPHORYLATION ON ENZYMACTIC ACTIVITY.
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -I- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y00624; CAA68663.1; -
CC PIR: A27224; A27224.
DR HSSP: P08799; 1MD.
DR InterPro: IPR000048; 1Q_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 2.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MSC; 1.
DR PROSITE: PS50096; IQ; 1.
KW Myosin; Coiled coil; Actin-binding; ATP-binding; Calmodulin-binding;
KW Methylation; Alkylation; Phosphorylation; Multigene family.
FT DOMAIN 1 789 MYOSIN HEAD-LIKE.

FT DOMAIN 790 819 IQ.
 FT DOMAIN 848 1509 COILED COIL (POTENTIAL).
 FT DOMAIN 848 1226 ALPHA-HELICAL TAIPIECE (S2).
 FT DOMAIN 1227 1252 HINGE.
 FT DOMAIN 1253 1509 LIGHT MEMBRANOSIN (LM).
 FT DOMAIN 1253 1482 ALPHA-HELICAL TAIPIECE (LM).
 FT DOMAIN 1483 1509 NONHELICAL TAIPIECE.
 FT NP_BIND 182 189 ATP.
 FT DOMAIN 660 682 ACTIN-BINDING.
 FT DOMAIN 766 780 ACTIN-BINDING.
 FT MOD_RES 133 133 METHYLATION (TR-1) (POTENTIAL).
 FT MOD_RES 700 700 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 1489 1489 PHOSPHORYLATION.
 FT MOD_RES 1494 1494 PHOSPHORYLATION.
 FT MOD_RES 1499 1499 PHOSPHORYLATION.
 SQ SEQUENCE 1509 AA; 171201 MW; 2CE49BE51173D17E CRC64;

Query Match 6.2%; Score 199; DB 1; Length 1509;

Best Local Similarity 20.4%; Pred. No. 0.33;

Matches 148; Conservative 124; Mismatches 276; Indels 176; Gaps 30;

QY 17 SNOSNMPIINGOLASSEKSTKASEASPSASSSSVSFLSAKKNALISRDALINK 76
 DB 740 STDVAKDLLEHLIA-----KEPTKYNK--DEVRFVTKLFFPSGOLAELELE----- 787
 QY 77 NSSPTDSLQLEASTSTSTVTRVAAKDYEAKSNFTAKSGLENAKTLAE----- 126
 DB 788 -----QAIKSMVYSIQAGARAFIARRYDKMRQGYSAKILQINIRAMELKNMAMYQLY 842
 QY 127 -----YETKADLMALQDMER-LANDSPSNHTEEVANNIKKALEAKDITDK- 173
 DB 843 VKARPLISQNFKEIDILKKQVDELEKELAAIKDAN--AKLDKEQLAEEDADKLEKO 899
 QY 174 ----LNKLVTLQONKSLTEVLKTTDSA--DQIPAINSOLEINKNSADQII--KDLER 223
 DB 900 LAALKIKILDEGEKADLEE-----DNALQKRYAGLEELQERTSANNILQKRLTEA 954
 QY 224 QNISYEAVLTNAGEVITKASSEAGIKLQALQSIYDAGDOSQA--VLQOQNNSPDNIA 281
 DB 955 EKGELKASTLEEERNRKALQEAKTVESENNELQDKYEDBAAHDSLKKKEEDLSRELRE 1014
 QY 282 TKE-LIDA---AEKYNELKQEHGTLD-----SPLVKKAEQISQAKQD 322
 DB 1015 TKALALAEINISFTLRKSLKNTERGADVDVATATKQLEKTKKSLLEELAQTBAQ 1074
 QY 323 IOETKPSGSDPIYGPSPGSAASAG-----SAAGALKSSNNSGRISLLDDVDNEMAAI 375
 DB 1075 LEEK-SGRE-----AASSKKQQLQGLEEDARSEVDSLKSLSAEKSLTKAKQON----- 1124
 QY 376 ALOGFRSMIEQFNVNNDATA--KEIQAMERQULTAMSDQLVGADGELPAEIQIKDALAQ 432
 DB 1125 -----RLDRLQLEDEERTVRANVDKOKKALBAKLTLELDQYATALDQCKNA-----AAAQ 1172
 QY 433 ALKQPSADGLATAMQVAFAPAAKYGGSSAGTACIVQNNVQIKYTAASSIS--SSSTAA 490
 DB 1173 A-----KTKTYQVDETKRRLKEAEASAA-----RLKERKNALDEVAQLADIDAE 1218
 QY 491 LSDGYSAYKTLNLSYSESRGVOA-----ISQTPAL 524
 DB 1219 RDSQAQQRKINTIRISLEQLELENAKPTGGASSEVKRLGELERLEELLTQAEARAAA 1278
 QY 525 SRSVSRSGIESQ--GRSADAQRAAETIVDSQTL-GDV-YSLQV-----LDS-- 569
 DB 1279 EKNIDKANLELEELRQADDAARNDKLVKDNKRLKADLEARQLEEDQDAKSHADSS 1338
 QY 570 --LMSTVSNQANQERIMOKLTASIKAPQFGVPANONSADSIQKRAQLERFVNGER 627
 DB 1339 RRLAELEELKKRYAKETSDOKRAQDQA-----NYQRENESLAKADRDSTERRNRAER 1392
 QY 628 SLAE 631
 DB 1393 QVRD 1396

RESULT 11
 ID MYH3_RAT STANDARD; PRT: 1940 AA.
 AC P12847;
 DT 01-OCT-1989 (rel. 12, Created)
 DT 01-OCT-1989 (rel. 12, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Myosin heavy chain, fast skeletal muscle, embryonic.
 GN MYH3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=101116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87060988; PubMed=3783701;
 RA Strehler E.E., Strehler-Page M.-A., Perriard J.C., Periasamy M.,
 RA Nadal-Ginard B.;
 RT "Complete nucleotide and encoded amino acid sequence of a mammalian
 RT myosin heavy chain gene. Evidence against intron-dependent evolution
 RT of the rod.";
 RL J. Mol. Biol. 190:291-317(1986).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X04267; CAA27817.1; -
 CC PIR: A24922; A24922.
 DR HSSP: P13538; 2MYS.
 DR InterPro: IPR000048; IQ_region.
 DR InterPro: IPR004009; Myosin_N.
 DR InterPro: IPR002928; Myosin_tail.
 DR InterPro: IPR002017; Spectrin.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF00612; IQ; 2.
 DR Pfam: PF01576; Myosin_tail; 1.
 DR Pfam: PF02736; Myosin_N; 1.
 DR PRINTS: PRO0193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR PROSITE: PS50096; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family.
 FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
 FT DOMAIN 782 811 IQ.
 FT DOMAIN 840 1933 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP (POTENTIAL).
 FT DOMAIN 656 678 ACTIN-BINDING.


```

FT DOMAIN 758 772 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 696 696 ALKYLATION (SH-1).
FT MOD_RES 706 706 ALKYLATION (SH-2).
FT MOD_RES 1940 AA: 223857 MW: BDD546A596E3A696 CRC64:
SQ SEQUENCE

Query Match
Best Local Similarity 21.5%; Score 199; DB 1; Length 1940;
Matches 155; Conservative 114; Mismatches 271; Indels 180; Gaps 33;

OY 8 MEGPSSVSSNOSMPT---INGQTA---SNSETKES-----TKASAPSP 48
DB 1232 LSSSVSSKSKANLEKICRTLEDQSEARGNKEFORSELTQKSRLOTEGELSP 1291
OY 49 ASSSVSWPFLSSAKNA---LISLADALINKSSPTDLSOLEASTSTSTYRVAAK 104
DB 1292 LEKESIVSOLSKQAFPOQIEELKROLEBKAKANALAHALQSSRHDCDLREYEE 1351
OY 105 DEAKSNFDTA--KSGLENKTLAEYETKMAADLMAALQDMERLANSPSNHTEEVN 162
DB 1352 QEGKALQALSKANSEVAQMRKYETD-----ATQ-----RTPELEAKK 1392
OY 163 ALEAQOTDIDKLVTLQNLQNSKLTVEYKTTDSADQIPAINSQ---LEINK---NSAD 215
DB 1393 KL-AQR-----LQ-----DSEQVEAVNAKCASEKTKORLOGEYE 1427
OY 216 QIIRKLERONISYEAVLITNAGEVIKASSEAGIKLGALQSYDAGQSOA---AVLQAO 272
DB 1428 DLMVDERANSLAALADKKORNFQVLAEMKTC-----BESQAELEALAKESR 1476
OY 273 NNSPDNIA---ATKELIDAETKVN---LKOEHGTLDSPYKKAEEQISOAKDIOEI 326
DB 1477 SLSTELFKLNVEEALDQLETKRENKMLEQEIADLT-----EQIAENGKSIHEL 1527
OY 327 KRS-----GSDPIVPGSGSASAGSAGALKSSNNSGRISLLDDVQDNEN-AAIALQ 378
DB 1528 EKSRKOMELEKADIQWALEEAEALHEHEEAKIL-----RIQLELTQVKSSEIDRKIAK 1580
OY 379 GFRSMTEOPNVNPPATKELQ--AMEAQLAMSD-----QLVGADGELPAEIQAIIDALA 431
DB 1581 D--EEIQKRNKQRTQRTMOGLDQAEVSRNEAIRLKKRMEDLNEIEIQSHANROA 1638
OY 432 QALKOPSSADGLATAMGOVAFAAKVGSGSAG-----TAGTQVMNVKQLYTA 478
DB 1639 EIKHK-----LRSVQQLKDQTLHDDALRGQEDLKEQLAIYERRANLQAEVEEL----- 1689
OY 479 FESSTSSSYAALSDGYSAYKTLNLSYSESRSQVQSAISQIANPALRSVSFR----- 530
DB 1690 -----RATLEQTERARKLAEOGLDLSNERVQ--LHTQNTSLIHTKKLETDLQQL 1738
OY 531 -SGIESGSSADAPRAAETIYRDSOTLGDVYSRLQVLDLSLSTIYSNQANOEEIMOKL 589
DB 1739 QSEVEDASDARNNAEKAKKAITDAMMAEELKEQ--DT--SAHLERRKKMLEQIVKDL 1794
OY 590 TASISAPQFGYPAYONSADSLQKFAAL--EREF--VDGERSLAESQENAFRKOPAFIOQV 647
DB 1795 QHRDEAEQL---ALKGKKQIQKLETRIRELEFELEGEQKNTESVKGKLYERRRVKEL 1851

RESULT 12
TALA_DICDI STANDARD: PRT: 2492 AA.
AC P54633:
DB 01-OCT-1996 (Rel. 34, Created)
DB 16-OCT-2001 (Rel. 40, Last sequence update)
DB 16-OCT-2001 (Rel. 40, Last annotation update)
DE Filopodin (Talin homolog).
GN TALA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=AX2;
RX MEDLINE=95213284; PubMed=7698984;
RA Kreitmeier M., Gerisch G., Heizer C., Mueller-Taubenberger A.;
RT "A tallin homologue of Dictyostelium rapidly assembles at the leading
RT edge of cells in response to chemottractant.";
RL J. Cell Biol. 129:179-188(1995).
RN [2]
RP REVISIONS TO 2076; 2080; 2186 AND 2230.
RA Mueller-Taubenberger A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RC -I- FUNCTION: ACTIN-BINDING PROTEIN THAT MAY BE INVOLVED IN THE
CC CONTROL OF CELL MOTILITY AND CHEMOTAXIS.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC; RAPIDLY ASSEMBLES AT THE
CC LEADING EDGE OF CELLS IN RESPONSE TO CHEMOTTRACTANT.
CC -I- SIMILARITY: CONTRAINS 1 BAND 4.1-LIKE DOMAIN.
CC -I- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C.ELEGANS ZK370.3.
CC PARTIAL, TO YEAST ROD1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: U14576; AAC46586.2; -
DB DictyDB; DD01121; TALA.
DB InterPro; IPR000299; Band_4.1.
DB InterPro; IPR002558; ILMEQ.
DB Pfam; PF00373; Band_41; 1.
DB Pfam; PF01608; ILMEQ; 1.
DB ProDom; PD011820; ILMEQ; 1.
DB SMART; SM00295; B41; 1.
DB SMART; SM00307; ILMEQ; 1.
DB PROSITE; PS00660; BAND_41_1; 1.
DB PROSITE; PS00661; BAND_41_2; 1.
DB PROSITE; PS00507; BAND_41_3; 1.
KW Structural protein; Cytoskeleton; Actin-binding.
FT DOMAIN 134 294 BAND 4.1-LIKE.
SQ SEQUENCE 2492 AA; 268850 MW; 4B82F6BE6E7E752 CRC64;

Query Match
Best Local Similarity 20.4%; Score 199; DB 1; Length 2492;
Matches 147; Conservative 113; Mismatches 279; Indels 180; Gaps 28;

OY 43 SEASPSSSVSSVFLSSAKNALIS---LRDAIINKNS---SPTDSLQLEASTSTST 95
DB 1831 STISNDIANYIDIAHVSSAIIATTSQKIRSIITSSNOYIVSGDIYHNHKVN-STDK 1889
OY 96 VTRVAADYDEAKSNFDTAKSGLENK--TLAEY--ETKMADLMAALQDMF----- 142
DB 1890 ANSSQAKVDSYRAVNDNITRFLQSLKQALIGEILSDAIDQIRRVISDLQYSLFAAG 1949
OY 143 RLANDSPSNHTEEVNNTKKALEAKDPTDKINKLVTLQNO----- 183
DB 1950 QLENDQSSOSTYNEVTKQOHLKNLQKDKTTLIVSSSQLVGSSRGTOEHLGSAATKV 2009
OY 184 -----NKSLEVLKTTDSADQIPAINSOLEINKNSADQIIRKLERONISYEAVLT 233
DB 2010 ANTVSSIVKTKANDIASVLADTTSSQDILSASAKLSSISSQAWLAKKDAQRFKKDTA-FR 2068
OY 234 NAGEVIKASSEAGIKLGALQS---IYDAG---DOSQAAVLAQOONNSPDNIATREL 285
DB 2069 SLGKSAEVAEAA---VCGFLTSVYTAISDAGKIGIELKSTIYQVANYHEKPTVLSNRDA 2125
OY 286 I-----DAETKVNLEKQEHGTLDSPYKKAEEQISOAQKDQIEIRPSSDPIIV 336
DB 2126 TAEITPQASRDIAKSSIT-EIYTSYSQDS-LVKSQSAVNSVQSFISNSK----- 2174
OY 337 GPSGSAASAGSAAALKS-----SNNSGRISLLDDV-----DNE 371
DB 2175 ---GVIALLGNGNDLKSQVLENVKQTTGDMALLQGVCKDDKNGSTSIADATRSISR 2231

```



```

RN (4)
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 777-836.
RX MEDLINE-96419133; PubMed-8805510;
RA Houdusse A., Cohen C.;
RT "Structure of the regulatory domain of scallop myosin at 2-A
RL resolution: implications for regulation.";
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55714; CAA39247.1; -.
DR PIR: S13557; S13557.
DR PIR: A40997; A40997.
DR PDB: ISCM: 30-APR-94.
DR PDB: 1WDC: 11-JUL-96.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head. 1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 1.
DR SMART: SM00015; MYSC; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
DR PROSITE: PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Alkylation; Calmodulin-binding; 3D-structure.
FT DOMAIN 1 777
FT DOMAIN 1 778 805
FT DOMAIN 1 836 1938
FT DOMAIN 1 836 1938
FT NR_BIND 176 183
FT MOD_RES 693 693
FT MOD_RES 703 703
FT HELIX 778 821
FT HELIX 822 823
FT TURN 825 833
FT TURN 834 835
SQ SEQUENCE 1938 AA; 222821 MW; ASCECE4127D1A4896 CRC64;
Query Match 6.1%; Score 197; DB 1; Length 1938;
Best Local Similarity 20.0%; Pred. No. 0.55;
Matches 148; Conservative 113; Mismatches 252; Indels 226; Gaps 29;
OY 5 HHMHSGPSSVSSNOSSNMPT-----INGQT-----ASNSERKEST-----K 41
DB 1334 HADMDAIRQLEEEQESKSDVOROLSKANNEIOQWRSKEPSEGANTEPELEQOKRLLGK 1393
OY 42 ASEASPSSASSVSSWFLSSAKNALISLRDALINKSSPTDSLQSEASTSTSTVTRVA 101
DB 1394 LSEADPTTAANAAKSALKAKRSRL-----QOELEDMSS-TEVVRANAASVQMK 1441
OY 102 KDYEAKSNFTAKSGLENKTLAEYETKMDLMAALQDMERLANSPPSNHTEEVNNTK 161

```

```

DB 1442 K-----QRAFD-----KTTAEQAKVNSIQSELENSOK-----ESRGSAELYRIK 1482
OY 162 KALEAQKPTIDKLNKLTVTLOQONKSLTEVLKTTDSADQIPALNSOLEINKNSADQI 221
DB 1483 ASIEEYQDSIG-----ALREKNKL-----ADELHDLTDQISEGGRSTHELDK 1527
OY 222 ERONISYEAVLITNAGEVINKASSEAGIKGLQALQSIYDADQSOAAVLQAO-----Q 275
DB 1528 RLEMEKE-----ELQALAEABGALQEGAEKAVMRQLEIATVARN 1569
OY 276 PNNIATKELID-----AAETVNEIKQETHGLTDSPLVKKAEQ----- 315
DB 1570 DKRIQKEEFPNTRRNHRALESQASLEAEKADAMRIKKLEQDINLEVALDAS 1629
OY 316 -----ISQAKQDIQEIIPS-----GSDIPIVSGSSA-----ASA 345
DB 1630 NRGKAEKMEKTVRRYQOQIREMOTSIIEEORQDEARESNMERRCTLMGSGVEELRA 1689
OY 346 GSAGALKSSNSGRISLLDDVD--NEMAA--IALQGRSMIEQFNVPNNPATARELO 401
DB 1690 EDAERARKASDME-----LADANDRVNELTSQVSSVQGGKRL-----GDINAM 1734
OY 402 EAQLTAMSDQVLGADGELPAEIQAIKDALQALQKPSADGLATAMGQVFAAKVGGSSA 461
DB 1735 QTDLEMHGELKCAD-----ERCKKAMADAAR--LADELRAEDDH-----S 1773
OY 462 GTAGTQVQNVKQLYTAPSTSSSSYAALSDQSYAKYKLNLSRSRSGVQSAISQTN 521
DB 1774 NOYEVKRNLESQVR-EFQIRLDEAEASSLKGGKKMIQLESRVHLEAELENE----- 1826
OY 522 PALSRVSRSGLSEOSGRSADASQRAEFTVRDSQTLGVYSRQLVDSIMSTIVSNPOAN 581
DB 1827 -----QRRAEFTQKNMKAPRLKE-----LAQDADEDRN 1857
OY 582 QE---EIMQKLTASISKAPQFGYPVAVNSADSLQNF-AAQLEREYVDSGLAESQENAF 637
DB 1858 QERLOELIDKLNKIKTFQRQVEAEIYAIINLAKYRKAKQHELEAEERADTADSTLQKF 1917
OY 638 R---KQAPFIQOVLNINAS 653
DB 1918 RAKSRSSVSVQSRSSVSA 1936
RESULT 15
TFM1_HUMAN
ID TFM1_HUMAN STANDARD: PRT; 1093 AA.
AC P82094;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TATA element modulatory factor (TMF).
GN TMF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC TISSUE=Cervical carcinoma;
RX MEDLINE-93028466; PubMed-1409643;
RA Garcia J.A., Ou S.-H.I., Wu F., Lusis A.J., Sparkes R.S., Gaynor R.B.;
RT "Cloning and chromosomal mapping of a human immunodeficiency virus 1
RT "TAR" element modulatory factor";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9372-9376(1992).
CC -1- FUNCTION: THIS PROTEIN BINDS THE HIV-1 TATA ELEMENT AND INHIBITS
CC TRANSCRIPTIONAL ACTIVATION BY THE TATA-BINDING PROTEIN (TBP).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L01042; AAD54608.1; -

DR TRANSFAC; T00835; -

DR Genew; HGNC:11870; TMF1.

DR MIM: 601126; -

KM Transcription regulation; DNA-binding; Repressor; Coiled coil.

FT DOMAIN 439 922 COILED COIL (POTENTIAL).

FT DOMAIN 984 1092 COILED COIL (POTENTIAL).

SQ SEQUENCE 1093 AA; 123170 MW; 26133EBE5F4677BE CRC64; .

Query Match 6.1%; Score 194; DB 1; Length 1093;

Best Local Similarity 19.2%; Pred. No. 0.36;

Matches 145; Conservative 141; Mismatches 279; Indels 190; Gaps 28;

QY 10 GSEPVSSNQSSNPPIINGQIASNS-ETKESTKSEASPSSASSSVSSFLSSAKNALIS 68

DB 169 SSPKTEGKHHEVYNKESDKMKVPTVSLKVESVIVDYKTMESISNTTOSTATKRIALE 228

QY 69 LRDALINKSSPTDLSLSQLEASTSTSTVTRAAKDYDE--ANSNFTPAKSGLENAKTAE 126

DB 229 PREQKHEDRQSNTPSPVPSTFSSGTSSTSDIEYLDHEYSIESSSASSRQETTDKSKSLHL 288

QY 127 YEKKADLMAA-----LQDMERLANSDPNNHTEEVNNIK-KALEAQKDTDKLTKLV 178

DB 289 MQTSFQLLSASACPEYNRLDDFOKLTESCCSDAFERIDFSVQSIDSR--SYSEINSD 346

QY 179 TLQNKSKSTLEV-----KTTDSAD-----QIPAINSOLEINKNSADQIIRK 220

DB 347 ELISGRKVALPPIVNSTSTPKTSKVESAEKSEEVNLTIVIPTEALMEBSGRATVY--N 404

QY 221 LERQNTSYPAVLTNAGEVIRKASSEAGIKGLQALQSTVDAQDSQA----- 265

DB 405 CEQPDILVSTSPINEGQTV-----LDKVAEQCEPAESQEPALSEKEDVCKTEFELN 455

QY 266 -----AVLQAQNNSPDNIAATKELIDAEKTKVDELKQEHGTLDSPL 308

DB 456 EKLEKREAOQLLSKKEKALLEEFDMLKDEMPVKE-----ESSSISLSLDEFT-----OR 506

QY 309 VKKAERQI-----SQAKDIOEIKPSGSDPIYVPGSSAASAGSAAATKSSNSGR 361

DB 507 IAEAEKKVOLACKERNAKKEIKNIKEE-----LATRLNSSET 544

QY 362 SLILDDVDNEMAIALQGRSMIEQFPVNNPATKELQAMEAQLTAMSDQLVGADGELPA 421

DB 545 ADLKREKDIQIRGLMEGRKLSKQ--LHNSNITIKIRAKDKENEMVAKINKKVELEE 602

QY 422 ELQATDALAQAALKOPRSAGGLATAMQVAFPAAKVGG--GSGTGTGYQVMNKQLYKTA 478

DB 603 ELQHLKQVL-----DGKEEVERKHRENIKKLINSWYERQEKDLGRLOVDMDL---- 649

QY 479 FSTSSSYAALSDGYSAVKTLSLY-----SESRSGVQSAISO--TAMPALSRVSRS 532

DB 650 --EKKRSTIOALID--SAVKELTLDHKANAADSEAFQALSRKAKAEKLSALEKKA 704

QY 533 ISSQ-----GRSADASQRAAEFTVDSQTL--GDVYSRLQ-----VLDSI 570

DB 705 EEARQOQETLAIQVGLRLALQRTQEAARKEDYLRRHEIGELQQRQLEAENRNQELSQSV 764

QY 571 MSIT-----VSNPQA-----NQEITQKLTASISKAPQFGTAVQNSANDSLQKRAQLERE 621

DB 765 SSTTRPLRLQRIENTLQATLQSTSSWEKLEKNS-----DRLGESQTLIAAVERE 814

QY 622 FVDGERSLA-----ESQENAFRKQPAFIOQVL 648

DB 815 RAATEELLANKIOMSMESONSILLROENSRFOQL 849

RESULT 16

MYH3_HUMAN

ID MYH3_HUMAN STANDARD: PRT: 1940 AA.

AC P11055; Q15492;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Myosin heavy chain, fast skeletal muscle, embryonic (Muscle embryonic

DE myosin heavy chain) (SMHCE).

GN MYH3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RT [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89263803; PubMed=2726495;

RA Eiler M.S., Steadman H.H., Sylvester J.E., Fertels S.H.,

RA Rubinstein N.A., Kelly A.M., Sarkar S.;

RT "Nucleotide sequence of full length human embryonic myosin heavy

RT chain cDNA.";

RT Nucleic Acids Res. 17:3591-3592(1989).

RL [2]

RP SEQUENCE OF 774-1940 FROM N.A.

RX MEDLINE=90033298; PubMed=2806546;

RA Eiler M.S., Steadman H.H., Sylvester J.E., Fertels S.H., Wu Q.-L.,

RA Raychowdhury M.K., Rubinstein N.A., Kelly A.M., Sarkar S.;

RT "Human embryonic myosin heavy chain cDNA. Interspecies sequence

RT conservation of the myosin rod, chromosomal locus and isoform

RT specific transcription of the gene.";

RL FEBS Lett. 256:21-28(1989).

RL [3]

RP SEQUENCE OF 856-1940 FROM N.A.

RC TISSUE=Skeletal muscle;

RX MEDLINE=90235862; PubMed=1691980;

RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goede H.W.,

RA Arnold H.H.;

RT "Identification of three developmentally controlled isoforms of human

RT myosin heavy chains.";

RL Eur. J. Biochem. 189:55-65(1990).

RL [4]

RP SEQUENCE OF 856-1940 FROM N.A.

RX MEDLINE=89366648; PubMed=2771643;

RA Karach-Mizrahi I., Travis M., Blau H., Lelwand L.A.;

RT "Expression and DNA sequence analysis of a human embryonic skeletal

RT muscle myosin heavy chain gene.";

RL Nucleic Acids Res. 17:6167-6179(1989).

RL [5]

CC - FUNCTION: MUSCLE CONTRACTION.

CC - SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC - SUBCELLULAR LOCATION: Thick filaments of the myofibrils.

CC - DEVELOPMENTAL STAGE: ABUNDANTLY PRESENT IN FETAL SKELETAL MUSCLE

CC AND NOT PRESENT OR BARELY DETECTABLE IN HEART AND ADULT SKELETAL

CC MUSCLE.

CC - DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC - PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY

CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.

CC - MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT

CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE

CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED

CC SUBFRAGMENT (S2).

CC - SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC - SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X13988; CAA32167.1; -

DR EMBL: X13100; CAA31492.1; -

DR EMBL: X51593; CAA35942.1; -

DR EMBL: X15696; CAA3731.1; -
 DR PIR: S04090; S04090.
 DR HSSP: P13538; 2MYS.
 DR Genew: HGNC:7573; MYH3.
 DR MIM: 160720; -
 DR InterPro: IPR00048; IQ_region.
 DR InterPro: IPR004009; Myosin_N.
 DR InterPro: IPR002928; Myosin_tail.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head.1.
 DR Pfam: PF00612; IQ; 2.
 DR Pfam: PF01576; Myosin_tail.1.
 DR Pfam: PF02736; Myosin_N.1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR Prodom: PD000355; myosin_head.1.
 DR SMART: SM00015; IQ.1.
 DR SMART: SM00242; MYSC.1.
 DR PROSITE: PS50096; IQ.1.
 KW Myosin; muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family.
 FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
 FT DOMAIN 782 811 IQ.
 FT NP_BIND 840 1933 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP (POTENTIAL).
 FT DOMAIN 656 678 ACTIN-BINDING.
 FT DOMAIN 758 772 ACTIN-BINDING.
 FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 696 696 ALKYLATION (SH-1).
 FT MOD_RES 706 706 ALKYLATION (SH-2).
 FT CONFLICT 1331 1331 A -> G (IN REF. 3).
 FT CONFLICT 1391 1392 K -> Q (IN REF. 1 AND 2).
 FT CONFLICT 1608 1609 SR -> RA (IN REF. 3).
 FT CONFLICT 1663 1664 RG -> QT (IN REF. 2).
 SO SEQUENCE 1940 AA; 224035 MW; 43CA58C6A4BA1253 CRC64;

Query Match 6.1%; Score 194; DB 1; Length 1940;
 Best Local Similarity 21.3%; Pred. No. 0.73;
 Matches 155; Conservative 113; Mismatches 265; Indels 194; Gaps 34;

OY 8 MESGEPSYSSNOSSMNP-----INGQIA-----SNSETKES-----TKASEASPS 48
 DB 1232 LSSSMESVSKSKANLEKICRTLEDDISEARGKNEIORSLSLTQKSRLOTEAGELSRQ 1291
 OY 49 ASSSVSSMFLSSAKNALI-----SLRDALINKNSPTDLSLEASTSTSTYTRVANDY 104
 DB 1292 LEEKESIYSQSRSKQAFTOOTEELKROLEENKAKMALAHALQSSRHDCDLRQOYEE 1351
 OY 105 DEAKSNFDTA--KSGLENAKTLAEYETKMDLALQDMERLANSNHTPEYNNIK 162
 DB 1352 QEGKRELQRAISKANSEVAQWRTKETD-----AIQ-----RTEELEAKK 1392
 OY 163 ALEAQKDTIDKLKLVTTQONKSLTEVLKTTDSADQIPAINSO---LEINK-----NSAD 215
 DB 1393 KL-AQR-----LQ-----DSEEOVEAVNAKASLEKTKORLOGEYE 1427
 OY 216 QIIRKLERONI-----SYEAVLT---NAGEVIKASSAGIKLQALQSIYDADQ 262
 DB 1478 DLMDVVERANSIALAALDKKORNFEDVLAEMWKCESSQELDASLKESRSL----- 1478
 OY 263 SQAAVLQAQOONNSPDNIATKELIDAETKYNE---LKOHTGLTDSPLVKAKEQISOA 319
 DB 1479 --STELFKLN-----AYEALDQLETVKREKNLEGEIDL-----QDAIEN 1520
 OY 320 QKDIOEIRPS-----GSDPIVPGSSASASAGSAGALKSSNNGRISLLDDVDNEM 372
 DB 1521 GRTHELKSRKQIELEKADIQILAEEAALHEEAKTL-----RIQLELTQVKSSEI 1573
 OY 373 -AAILQGFRRMIEQFNVNPNPATKELO-AMEAQLTAMSD-----OLVAGDGEPAEIQ 424
 DB 1574 DRKTIKED--EIEQLKRRYQTFVTETMOSALDAEVRSRNEAIRLKKMGDLNIEIQIS 1631
 OY 425 AIKDALQALKOPASDGLATAMGOVAFAAKVGGGSAG-----TAGTVOMNV 471

DB 1632 HANROAETLKH-----LRVSGQLKDTQULHDDALRCQEDLKEQLATVERRANLQAEV 1686
 OY 472 KQLYTAFTSSSTSSSYAAALSDGYSAYTTLNSLYSESSGVSQSAISQTAANPALSNSVR- 530
 DB 1687 EEL-----RATLEOTERARKLAEDLSDSNERVO--LHTRQNTSLHTKKKL 1731
 OY 531 -----SGIESQGRSADASORAEITVRSQTLGDVYSRLQVLDLSLMTSTVSPQNO 582
 DB 1732 ETDLMQLOSEVEDASRDARNKEKAKKAITDAMMAEELKEQ--DT--SAHLEMKRNL 1787
 OY 583 EIMQKLTASISKAPQFQYPAVQNSADSLQKFAQL-REF-VDSERSLAESQENAFKQ 640
 DB 1788 EGTVDLQHRDEAEQL--ALKGKKQIQKLETRIREFLEFELEEQKKNESVGLKRY 1844
 OY 641 PAFIOOV 647
 DB 1845 ERRVKEL 1851
 RESULT 17
 MISA_DROME STANDARD: PRT; 1962 AA.
 AC P05661;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Myosin heavy chain, muscle.
 GN MHC.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN=Canton-S; TISSUE=Pupae;
 RX MEDLINE=89384556; PubMed=2506434;
 RA George E.L., Ober M.B., Emerson C.P., Jr.;
 RT "Functional domains of the Drosophila melanogaster muscle myosin heavy-chain gene are encoded by alternatively spliced exons.";
 RL Mol. Cell. Biol. 9:2957-2974(1989).
 RN [2]
 RP SEQUENCE OF 1-264 FROM N.A.
 RX MEDLINE=87280141; PubMed=3038896;
 RA Wassenberg D.R., II, Kronert W.A., O'Donnell P.T., Bernstein S.I.;
 RT "Analysis of the 5' end of the Drosophila muscle myosin heavy chain gene. Alternatively spliced transcripts initiate at a single site and intron locations are conserved compared to myosin genes of other organisms.";
 RL J Biol. Chem. 262:10741-10747(1987).
 RN [3]
 RP SEQUENCE OF 333-614 FROM N.A.
 RC STRAIN=Canton-S; TISSUE=Embryonic muscle;
 RX MEDLINE=91330870; PubMed=1907912;
 RA Kronert W.A., Edwards K.A., Roche E.S., Wells L., Bernstein S.I.;
 RT "Muscle-specific accumulation of Drosophila myosin heavy chains: a splicing mutation in an alternative exon results in an isoform substitution.";
 RL EMBO J. 10:2479-2488(1991).
 CC -I- FUNCTION: MUSCLE CONTRACTION.
 CC -I- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2)
 CC -I- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -I- ALTERNATIVE PRODUCTS: At least 15 isoforms: AAAA (shown here), BBBBA, BABDB, 3b, 7b, 7c, 7d, 9b, 9c, 11b, 11c, 11d, 11e, 15b and 18; may be produced by alternative splicing. Exons 3, 7, 9, 11 and 15 are mutually exclusive splicing exons and exon 18 is included or excluded
 CC -I- TISSUE SPECIFICITY: Expressed in larval and adult muscles. Isoforms containing exon 9a are expressed in indirect flight muscles, exons 9a and 9b are expressed in jump muscles, exons 9b

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RC MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles D., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltham J., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Murphy L., Niblett D., Odell C.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grynponez B.,
 RA Welfens I., Volckaert E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard J., Tallada V.A., Garçon A., Thode G.,
 RA Dague R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Shkurtel L., Lowe T., McCombie W.R., Paulsen I., Petchkin J.,
 RA Shkurtel L., Lowe T., McCombie W.R., Paulsen I., Petchkin J.,
 RA "The genome sequence of Schizosaccharomyces pombe";
 RA Nature 415:871-880(2002).
 RN [2]
 RP SEQUENCE OF 495-1727 FROM N.A. AND CHARACTERIZATION.
 RC STRAIN=972;
 RC MEDLINE=20123449; PubMed=10660053;
 RA Jimenez M., Petit T., Gancedo C., Goday C.;
 RT "The aml1 gene from Schizosaccharomyces pombe encodes a coiled-coil
 RT protein that associates with the medial region during mitosis";
 RL Mol. Gen. Genet. 262:921-930(2000).
 RN [3]
 RP SEQUENCE OF 644-834 FROM N.A.
 RC STRAIN=968 h90;
 RC MEDLINE=20223868; PubMed=10759889;
 RA Ding D.O., Tomlita Y., Yamamoto A., Chikashige Y., Harguchi T.,
 RA Hiroka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 RT fission yeast cells by the use of a GFP-fusion genomic DNA library";
 RL Genes Cells 5:169-190(2000).
 CC -1- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
 CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
 CC CYTOKINESIS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AL133357; CAB62414.1; -
 DR EMBL: AF010473; BAB65416.1; ALT_INT.
 DR EMBL: AB028012; BAB87316.1; -
 KW Coiled coil.
 FT DOMAIN 57 361 COILED COIL (POTENTIAL).
 FT DOMAIN 443 463 COILED COIL (POTENTIAL).
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1106 COILED COIL (POTENTIAL).
 FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).

FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
 Query Match 6.0%; Score 192.5; DB 1; Length 1727;
 Best Local Similarity 19.6%; Pred. No. 0.73;
 Matches 140; Conservative 131; Mismatches 286; Indels 157; Gaps 29;
 QY 13 EASVSNQSSNPFINGQINLSN-SEPKESKASEASPSASSVSSWMSAKNALISLSD 71
 DB 101 ESKITLLENQNRKFDALITENOSLRANSELOQSTIASQ-----LSIADQI---E 150
 QY 72 AILNKS-----SPTDSLQLE-----ASTSTVTYVAKDYDEAKSNPTAKS 116
 DB 151 ALQNSHSLGEOVSQAHQALSDIEERKKQHPMASSSSRYKEIILVQEKALVSLASLOS 210
 QY 117 G-----LE-NAKTLAEETKMDALQOM-ELANSDSNNHTEVNNIKKALEQ 167
 DB 211 DSKVCEKLEVSRRQVQDLKKLAGLAQONTLENERIQLEQKRNSYSSDGINISKLETD 270
 QY 168 KPTIDKRLVTLQONKSLTEVLKTTDSADQIPAINSQL-----EINKNSAD 215
 DB 271 PSIKLEEEVEYQKRLTALME-SKSSSELQSEVAALQEKLTQOQSLYNNVTELNKKQ 329
 QY 216 QIIKD--LERQNTSYEAVLTLNAGEVIKASSSEAGIKGLQALQSLVDAGDSQAIVLQAOQN 273
 DB 330 LITSESLRELQEKYVSSEL-QVYKRNKNTSVSAGVLFSPL-----AKLSAVON 381
 QY 274 NSPDNIAAKRELID--AAETYNELKQHTGLTD--SPLYKKAERQISOAQDIOIKRS 329
 DB 382 ---PEPSFKVYSDNNKLOQKVSLSKLQDLRLNFKSSFCQVKQKIPVAKQORSEIVRN 438
 QY 330 GSDIPVPGSGSAAAGSAAAGLAKSSNNSGRISLILDDVDNEMAAITAIQFMSIEQFNV 389
 DB 439 NIYMNFLSES-----LETSNNN-----LTQVQAEILSTKMN----- 469
 QY 390 NNPATKELQMEAOITLAMSQDLVGADGL-----PAETQAIKDALQALQKOPS 438
 DB 470 -QECALQTLASTQCSDSLSREVICLMAELDLNETKSRHNPATVQVLADEVAQ---NPS 525
 QY 439 A-----DELATAMGOVAFPAAKVGGSGAGT--AGYOVMMNKOLYKTAFSST 482
 DB 526 TASETVLNKELANFSSIKKAVSKTLELRKVRALQEDVETQOTVOYQISNAVKE--NSN 583
 QY 483 SSSSYAALSDGYSAYKTLNSLYSESRSQVQASISQATNPALRSVRSRSGISQGRSADA 542
 DB 584 TLSEQIKNLESELNKSINKNESLILNRLKELANRSRSLISHNSAGNIDKMSIDE 643
 QY 543 SORPAETIVRSQTLGDVTSRLQVLDLSMSTIVSNPQANQDEIMQKLTASISKAPOFGTP 602
 DB 644 STRELE-----KNEVEYRNEMTALQIESLSKRNDLSEMA-IRK----- 682
 QY 603 AVQNSADSLQKFAQLEREYVQGERSLASQENAFKQAPAFIQVIVNLSLFS 656
 DB 683 ELENS-----KITQOQLSTRLTN---ANNVDYAEFKKELKSLINQNLDTIIS 727
 RESULT 20
 MYH4_RABIT
 ID MYH4_RABIT STANDARD; PRT; 1938 AA.
 AC Q28641;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE 1500 heavy chain, skeletal muscle, juvenile.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=New Zealand white; TISSUE=Skeletal muscle;
 RA Maeda K., Hostinova E., Roesch-Kleinhauf A., Schuster H., Gasperik J.,


```

RA Wittinghofer A.;
RT "Isolation, sequencing of myosin heavy chain cDNA from rabbit
RT skeletal muscle and a novel cosynthesis of S-1 fragment with the
RT essential and regulatory light chains."
RL Submitted (Aug-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATOR LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32574; AAA74199.1; -.
DR HSSP: P13538; 2MYS.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; Myosin_head.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF00612; IQ.2.
DR Pfam: PF01576; Myosin_tail.1.
DR Pfam: PF02736; Myosin_N.1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head.1.
DR SMART: SM00015; IQ.1.
DR SMART: SM00242; MYSC.1.
DR PROSITE: PS50096; IQ.1.
DR MYOSIN: Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 783 MYOSIN HEAD-LIKE.
FT DOMAIN 842 813 IQ.
FT NP_BIND 179 186 COILED COIL (POTENTIAL).
FT DOMAIN 658 680 ATP (POTENTIAL).
FT DOMAIN 760 774 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 35 35 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (MONO-) (BY SIMILARITY).
FT MOD_RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 698 698 METHYLATION (SH-1) (BY SIMILARITY).
FT MOD_RES 708 708 ALKYLATION (SH-2) (BY SIMILARITY).
SQ SEQUENCE 1938 AA; 223064 MW; D8A8A2EC5B182626 CRC64;

Query Match 6.08; Score 192; DB 1; Length 1938;
Best Local Similarity 19.68; Pred. No. 0.88;
Matches 111; Conservative 107; Mismatches 227; Indels 122; Gaps 22;

```

```

DB 949 ECSEELKDDIDDELTLRAV-EKEKHATEKNVNLTEPMAGLDETIYAKLT-----KEKKALQ 1003
QY 269 QAOQNSNDPNTAATTELLDAETKYKNELKQEHTEGLTDSPLVYKAEQISOAKQDQELKP 328
DB 1004 EAHQOTLDD-----LQAEEDKYNLTAKT-----KLEQOYDDLEGSLEQEK 1046
QY 329 SCSDDPIVPGSASAGSAGALKSSNNSGRISLLDDVDNEMAAIALOGFRSMIEDQFN 388
DB 1047 IRMDL-----EAKKKLEBDLKLAEOSTM-----DIENDKQDLDEKLRKKEFEFSN 1092
QY 389 VNNPATAKELAMEAOLTA MSPDOLVAGDELPAAEIOAIKDALAOLKOPSDGLTAMQ 448
DB 1093 LQ--SKIDEQOLAMQLOKIKTEILOARIELEEIEARASRAKAKQRS--DISRELEE 1148
QY 449 VAFAAKVGSGSAGTAGYQNMVVKQLYTFASTSSSSYAAALSGYSAKYTLNSLYSES 508
DB 1149 ISERLLEAGG--ATSAQIEKKR-----EAEFQKM 1177
QY 509 RSGVQSA-ISQTPANPALSRSVSRSGIESGRSADASORAETIYVDSOTLGQVYSRLQVL 567
DB 1178 RROLEATLQHEATATATLRKXADSVAGELQIDMLQVKQKLEKSEL-----KMEI- 1231
QY 568 DSLMSTI--VSNPQANOEIMOKLTASIS--KAPQGYPAVQNSADSLOKFAOALEREFTV 623
DB 1232 DDLASNMETVSXAKGNLEKMCRTLEDQVSELKTRKEEH---QRLINDLSAQRARLQTESG 1288
QY 624 DGERSLAESQ---ENAFRKQPAFTQOV 647
DB 1289 EFSRQLDKESLIVSOLSRSKQAFTOOI 1315

RESULT 21
ANT1.ONCVO STANDARD; PRT; 2022 AA.
ID ANT1.ONCVO STANDARD; PRT; 2022 AA.
AC P21249;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Major antigen.
GN OVT1.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Splirurida; Filarioidae;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95287898; Pubmed=7770081;
RA Tritreaprapab S., Riche T.L., Tuan R.S., Shepley K.J., Dلمان J.D.,
RT Neubert T.A., Scott A.L.;
RT "Molecular cloning of a gene expressed during early embryonic
RT development in Onchocerca volvulus."
RL Mol. Biochem. Parasitol. 69:161-171(1995).
RN [2]
RP SEQUENCE OF 733-866 FROM N.A.
RX MEDLINE=89127417; Pubmed=2464764;
RA Donelson J.E., Duke B.O.L., Moser D., Zeng W., Etroundu N.E.,
RA Lucius R., Renz A., Karam M., Flores G.Z.;
RT "Construction of Onchocerca volvulus cDNA libraries and partial
RT characterization of the cDNA for a major antigen."
RL Mol. Biochem. Parasitol. 31:241-250(1988).
CC -1- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U12681; AAA80009.1; -.
DR EMBL: J03995; AAA29412.1; -.

```



```

FT CONFLICT 250 250 L -> Q (IN REF. 2)
FT CONFLICT 272 272 G -> D (IN REF. 2)
FT CONFLICT 611 611 MISSING (IN REF. 2)
FT CONFLICT 1494 1589 MISSING (IN REF. 2)
FT CONFLICT 1611 1611 V -> A (IN REF. 2)
FT CONFLICT 1811 1811 V -> L (IN REF. 2)
FT CONFLICT 2242 2243 ER -> DG (IN REF. 3)
FT CONFLICT 2335 2335 L -> Q (IN REF. 3)
FT CONFLICT 2492 2492 D -> N (IN REF. 2)
FT CONFLICT 2545 2561 ENERVAALHNDDEACK -> SSMREMOFCIMTKRPVS
      (IN REF. 3)
SQ SEQUENCE 3210 AA; 367589 MW; 11D83324960DE4334 CRC64;

Query Match 5.9%; Score 189.5; DB 1; Length 3210;
Best Local Similarity 20.6%; Pred. No. 2.1;
Matches 164; Conservative 126; Mismatches 292; Indels 213; Gaps 36;

QY 13 EGVSSNOSMNPINGQINSNETKSTKASESPASSSVSMPLSAAKNALISLRQA 72
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 949 ETLSEKEMSSITIS---LNKREIELTQENGTLKEKINS---LNQEKMLTQKSES 999
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 73 IINKNSPPTDSIQ-----LEASTSTSTVTVAADY-----DE 106
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1000 FANYIDERKKSISELSDYKOEKILLQCEETGNAYEDLSQYKRAQEKNSKLECLNE 1059
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 107 AKSNEDTAASGLE---AKTLAETETMA-----DLMAALQMERIANS--DPSN 151
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1060 CYSLCENRKNELQLEAKFAKEHOEFLTKLFAEERNQMLMELETVQALSEMMDNQN 1119
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 152 NHTEEVNNIKKALAEAKDTIDKLNKLYT--LQONKSKLETVLKITTSADQI--PAINS-- 205
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1120 NKRSEAGLKOELMTELEKQNKQKEVNDLLQENQMLKMTKHCQMLESEPIKNSYK 1179
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 206 -----QLEINKNSADQIIRDLERONISYEAVALTNAGEVTKASEAGIKTG 250
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1180 EKESEKNOCNKRPQMDLEKREISLD---SYNAQVQLAEMLRNKRLKLOESEKECEQ 1235
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 251 QALQSTVADAGDSQAVALQAOQNSPDMIAATKEI--IDAAE----- 290
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1236 HELQTRGDELSTNLQDMQSE-----ISGLKCEIDAEKYSIGPHLESTSQNDNAH 1289
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 291 -----TKVNEK-----QERTGLTDS-----PLVKAKEQISOAQKIOEI 326
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1290 QCSLQTTNNKMLEKEICETILOAEKYELVTELENDSSRECIATIRKKAEEG---KLNLE 1346
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 327 KPSGSDIPVIG-----PSGSAAASAGS-----AAGALKSSNNSGRISILLDPVNEMA 374
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1347 KILNDSGLHGEIVEDIGEGEBOHPNQHPSLAPLDESNSYEHILTSDKEVQKHFAE 1406
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 375 -----TALQGFPSMTIEQFVNNPATAKELQ---AMEAQLTAMSDQLVGADGELPAEIO 424
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1407 LOEKFELTQSEKILHDOHCQMSKSELOQTVYDSLAKENLVLTNLRFQGLVYEMQL 1466
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 425 AIKDLAOLAKO---PSAAGLATAMQVAPAAA---KYVGGG-----AGTAGVQANVQ 473
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1467 GLEEGVLPPLSSCPVDSSSL--SLSDSSFYRLLEQOTQDMLSLNLDEAVSANQCSYDE 1525
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 474 LYKTAFSSRSSSYAALADGYSAVYKTLNSLYSESRGVSASISQANPALRSYV---SSS 531
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1526 VFCSSLTQTVYDSLAKENLVLTNLRFQGLVYEMQLGIEGLYV---PSLSSCYPPDS 1581
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 532 GIESGRASADASQRAETIVRDSQTLGDVYSRLQVLDSLMSTIVSPQAO----- 582
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1582 SLSSLDSSFFYRLLEQT-----GDM-----SLLSNLEGVANSQCSVDEVFS 1625
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 583 -----EIMQKTLASISKAPQFGYPAVQNSAD-----SLQFPAOLE-----REFYDGER 627
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1626 SLQEEFLTKRPPS---AAKGVLEESLCEVYRQSLTELEKMEGQITMKKEIOLEQ 1682
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 628 SL-AESQE--NAFRQ 640
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1683 LLSSEKQELDCLRKQ 1697
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 26
YACB_CAEEL
ID YMB_CAEEL STANDARD; PRT; 918 AA.
AC P34487;
DT 01-FEB-1994 (Rel. 28. Created)
DT 01-FEB-1995 (Rel. 31. Last sequence update)
DT 15-JUN-2002 (Rel. 41. Last annotation update)
DE Hypothetical 96.6 kDa protein F59B2.12 in Chromosome III.
GN F59B2.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berts M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Lalister N.,
RA Latelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopra A., Saunders E., Showkeen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; 211505; CAJ7581.1; -
DR PIR; S31132; S31132.
DR WormPep; F59B2.12; CE01024.
KW Hypothetical protein.
SQ SEQUENCE 918 AA; 96560 MW; E464F86B14945DE CRC64;

Query Match 5.9%; Score 188.5; DB 1; Length 918;
Best Local Similarity 20.2%; Pred. No. 0.49;
Matches 146; Conservative 115; Mismatches 263; Indels 197; Gaps 34;

QY 3 HHHHMGSPRESVSNOSMNP-INGQ-IASNETKE-----STKASESPASSSVSSM 56
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 74 HGHHSOSS---SSNTHSLTYVGADGKNITENSEKKDYKNSKYDEANENYTKIKSADG 130
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 57 SFLSSAKNALISLRALINKNSPPTDSLSQLEASTSTSTVTVAADKYDEAKSNEDTAKS 116
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 131 SVLETGKS-----KNGTMLSNNKINN-----HNRSSDASAY 149
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 117 GLE-NAKTLAEYTKADLMALQMERLANDSPNNHTEEVNNIKKALEAKDTIDKLN 175
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 150 GLEKSSKTYAD-----KNGTMLSNNKINN-----QSRSAALDEGN 186
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 176 KLYTLQNNKSL---TEVLKTTDS---ADQIPAINSOLEI-----NKN 212
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 187 ERYNQONADGTFLRNNTGKNTDEHLSHNVLDENQMSIGAGCTSNITNRKGSVDSDIN 246
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 213 SADQIIKDLERONISYEAVALTNAGEVTKASEAGIKLGOALQSTVADGQSOAAVLAQAO 272
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 247 AASDAHSNESIDAGCNKKSQVYKKAASAGSNADFEENLSEIKNAQDGSMSNSTGNFN 306
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 273 NNSPDNIAA---TKELDAETKYNELKQERTGLTDSPLVYKAEQISOAKDIOEIKP 328
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Db 307 NTSYDKATAEEYMSKKNNADGT---SMSEASHAG-SNMSKTKNSASQ-----350
QY 329 SGSDIPYVPGSGSAASAGSAGALKSSNNSGRI-SLLDDVDNEMAIALO---GFRSM 3833
Db 351 -SSDLMSHPNG-----IKSHSTSKTKDNVALDENQASGISISDQICKNGRSL 3988
QY 384 IE-----QNVNVPATAKELQAMENQOLTMSQLYGADGELPAE-----IQATKDA 4298
Db 399 NESSIESCRKKEBSRNNVTADTILDSVDAGTYSSSHKSASAGTSLDENHNKTHALQASVDE 4588
QY 430 LQAOLKOPESADGLATAMQOVFAAFAKVGGSAGTAGTV-QMNVKOLYKTFASSTSSSYA 4888
Db 459 HGN-WKNNISIDGSTRNKKTGFEFGSEMSASIKNADGTMSOVNFK-----NDTRNRYTE 5100
QY 489 A---ALSDGYSAKYTKLNSLYESRSYGVOAISOT---ANPAL-SRVSRSIGIESQGRSAD 5411
Db 511 AEKSALEKRNHE--KNSDGTFEDESKGSNSRYNRTDGGSNLAVGSIYVGKGGVSSNETIAS 5688
QY 542 AS-----QRAAE-----TIVDSQTLDDVYSRLQVLDLSMTIYSNPQAN 5811
Db 569 SNAFPTSABESNQDHLHQKTANCTETTHAADSQVAAASANAANSLDTSMSAA--DAKGN 6266
QY 582 QEEIMOKLTASISKAPQFGYPAVONASDSLQKFAQLER-EFYDGGERS-LAESQENAFRK 6399
Db 627 -----KYDKTSSQAAD-SHDAISASSDVOAKYKHAHDBEESISNDSSNQTASAEHNDSSK 6799
QY 640 Q 640
Db 680 Q 680

```

RESULT 27	
YIQ9_YEAST	
ID	YIQ9_YEAST
	STANDARD;
	PRT;
	995 AA

DT		01-FEB-1995 (Rel. 31, Created)
DT		01-FEB-1995 (Rel. 31, Last sequence update)
DT		15-JUN-2002 (Rel. 41, Last annotation update)
DE		Hypothetical 99.7 kDa protein in SDL 5' region precursor.
GN		YIL169C OR YI19402.07C.
OS		Saccharomyces cerevisiae (Baker's yeast).
OC		Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX		Saccharomycetales; Saccharomycetaceae; Saccharomyces.
RN		[1]
RP		SEQUENCE FROM N.A.
RC		STRAIN-S288C / AB972;
RA		Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA		Churchill C.M., Connor R., Cooney T., Dear S., Devlin K., Fraser A.,
RA		Gentles S., Hamlyn N., Hornell T.S., Hunt S., Jagels K., Jones M.,
RA		Louis E., Lyne G., Moule S., Moule T., Odell C., Pearson D.,
RA		Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RL		Walsh S.V., Whitehead S.;
CC		Submitted (DEC-1994) to the EMBL/Genbank/DDJ databases.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC		the European Bioinformatics Institute. There are no restrictions on ways
CC		use by non-profit institutions as long as its content is in no way com-
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (See http://www.isb-sib.ch/announce/ ,
CC		or send an email to license@sib-ch)
CC		-----
DR		EMBL; Z46921; CAAB87023.1; .
DR		SGD; S0001431; YIL169C.
DR		InterPro; IPR004089; Chmtaxis_transd.
DR		InterPro; IPR000727; T_SNAARE.
DR		PROSITE; PS50192; T_SNAARE; UNKNOWN_1.
KW		Hypothetical protein: Signal.
FT	1	SIGNAL
FT	23	POTENTIAL.
FT	95	HYPOTHETICAL PROTEIN YIL169C.
FT	92	T-SNAARE COILED-COIL HOMOLOG (POTENTIAL).
FT	154	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	28	CARBOHYD

FT	CARBOHYD	35	35	N-LINKED (GLCNAC. . .)	(POTENTIAL)
FT	CARBOHYD	468	468	N-LINKED (GLCNAC. . .)	(POTENTIAL)
FT	CARBOHYD	664	664	N-LINKED (GLCNAC. . .)	(POTENTIAL)
SQ	SEQUENCE	995 AA;	99735 MW;	P63E287A03F137EC	CR664;

Query Match	5.98	Score 188.5	DB 1	Length 995
Best Local Similarity	20.98	Pred. No. 0.54		
Matches 142	Conservative 119	Mismatches 274	Indels 143	Gaps 27

Qy	10	SGPEPVSNNOSNMNI-----INCQIANS-----ETMESTR--ASEMSPS	48
Db	27	: : : : : : : : : :	86
Qy	49	ASSVSSMFFLSAANALISLRDAIILKNSEPTD--SLSOLEASTS--STSTVRVAKD	103
Db	87	: : : : : : : : : : : : : : : :	144
Qy	104	YDEAKSNFDTRKSGLENAKTLAEYETKMAIDMAALQDMERLANSDDPNHNHTEEVNNIKKA	163
Db	145	SDVSSVSQSSSSASADVSSVSQSSASASDVSSSV--SQSASSTSDVSSVSQSSSSASDV	203
Qy	164	LEAQKOTIDKLKLVTTLDONQKSLTEVLKTTDSADQIIPALNSOLEIKNSADQI IKDLER	223
Db	204	SSSVSSQSSSSASADVSSVSQSSAS-----STSDVSSVSQSSASSTSDVSSSGSOVSASG	258
Qy	224	QNISVEAULTNAGEVIRKASSSAGIKIKQALQSLYDADQDQANVLQAOQNNSPD-----	277
Db	239	SSSSSEPOSTSS-----ASTASGANTSNSLSSTSSA--SSASATASNSLSSSDGTTILP	310
Qy	278	-----NIAATKELIDAEFTKKNELKOEHTGLT-----DSPLKKAKEQISQA	319
Db	311	TTTTISGDLTLTGKVIATGCVYVAAAGAKTILLDGKISFSPADLKRYGDLVYKSKRETYPGT	370
Qy	320	QKDIQEIKPSSGSDPIPVYSGSSASAGSAAACAL--KSSNSGRISILL-----	365
Db	371	EFDI-----SGENFDVTCNFAEBSAATSAISYGFPTSPDNGSDISLSKSKKEGVTF	425
Qy	366	DDVDNEMA-----AIALDGFMSME-----QFNVNNPATKELQAMEAQLT--	406
Db	426	SPYSNSGAFSFSNAILNGSVSGQLRDDEGVSNNGEIMLDNGSTVIYIEPVSGKGTVN	485
Qy	407	-----AMSDQJVGADGLPEAIOAIKALQALQKPSADBLATAMQVAFMAA	434
Db	486	IISGNLYAHYDPTFGQVYFKEG--GVLAVDPTETNAPPIPVG--YTGKNOIATTAQ	540
Qy	455	KVGGSGAGTAG--TYQNMVKOL--YKTFASSTSSSYAALSDGYSAVYKTLNSLSESR	509
Db	541	ITALSYDQTTGLVLTATGCKNRPFAITGTGSSSPFSYSEGI FACAATAYILYNTGCVATS-	599
Qy	510	SGVOSAIISQTNAPALSRSYRS--GIESQGRASADASQRAETIYVRDSQTLG-----D	559
Db	600	-----AASSSTASGA-SASVTCSTSFGASVYGTASTSFGASVYGTASTSFGASVYGTST	654
Qy	560	VY-SRLQVLDLSLMTIYS	576
Db	655	VYTTTLDYVNAFSTVYVS	672
RESULT 28			
MISN_DROME STANDARD; PRT; 2017 AA.			
AC	Q99323;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Myosin heavy chain, non-muscle (Zipper protein) (Myosin II).		
GN	ZIP.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;		
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
NCBI	taxid=7227;		
LN	[1]		


```

CC -----
DR EMBL; D89992; BAA2069.1; -
DR EMBL; D50476; BAA09069.1; -
DR EMBL; D43700; BAA07802.1; -
DR HSSP; P13538; 2MYS.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head.1.
DR Pfam; PF00612; IQ.1.
DR Pfam; PF01576; Myosin_tail.1.
DR Pfam; PF02736; Myosin_N.1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head.1.
DR SMART; SM00242; MYSC.1.
DR PROSITE; PS50096; IQ.1.
DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Calmodulin-binding;
KW Multigene family.
FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
FT DOMAIN 782 811 IQ.
FT DOMAIN 812 839 HINGE.
FT DOMAIN 840 1935 COILED COIL (POTENTIAL).
FT NP_BIND 178 185 ATP (POTENTIAL).
FT DOMAIN 659 681 ACTIN-BINDING.
FT DOMAIN 761 775 ACTIN-BINDING.
FT MOD_RES 129 129 METHYLATION (TR1-) (POTENTIAL).
FT MOD_RES 699 699 ALKYLATION (SH-1).
FT MOD_RES 709 709 ALKYLATION (SH-2).
SQ SEQUENCE 1935 AA; 221599 MW; 9A1244B67D63C83B CRC64;

```

Query Match 5.9%; Score 187.5; DB 1; Length 1935;
 Best local similarity 22.0%; Pred. No. 1.3;
 Matches 158; Conservative 89; Mismatches 265; Indels 205; Gaps 31;

```

QY 14 SVSSNOSNNPIINGOIASNSETKESTKASEASPSASSVSSWFLSSAKNALISLSDAT 73
DB 852 AMKENYEKKEDLTAKLAKKELEEKVMS-----LIQEKNDL-QLVTA 894
QY 74 LKNSSEPTD-----SLQLEASTSTSTVTVAAKDYDEAKSNDPTAKSGLENNKTLLA 125
DB 895 ESENLSDAERCGLIKTSKIQLEAKLKE--TNERLEDEEINAEITAKKRKLEDECSL 951
QY 126 EYETKNADLMAALQDMERLANSDPNNHTE-----VNNIKKAL-EAOKDTIDKL 174
DB 952 KRIDDELLETLAVKEKATEKNKVLNPEMASQDESTAKLTKKEKALQEAHQITLDDL 1011
QY 175 NKLVTLONONKSLTEVLKTTDSADQIPAINSOLEINKNSADQITIKLERONISYEAVLN 234
DB 1012 Q-----AEEDKVTTLTKAKTKLEQOVDDLEGSLEOEK---KLMDLERAKRKLEGLKL 1062
QY 235 AGEVIT-----KASSEAGTK-----LGQALQSTVDAGDOSQAVALQAOONNSDNTAATK 283
DB 1063 AQSISIDLENKQSDKIKKKDFEISQLSKIED--EQLGALQOKK-----IK 1110
QY 284 ELIDAEFTYVNLKQHTGLTDSPLVKKAEQISQAOKDOETKPGSGDIPYVPGSGSAA 343
DB 1111 EL-----QARIELELEIE--AERAARAYEKQADSLRELEI-----SERLE 1152
QY 344 SAGSAAAGALKSSNNSGRISILLDDVDNEAATAIALQGRSMIEFNNPNPATKELQAMEA 403
DB 1153 EAGGATATA-----QIEK-----NKKREAFQKMRRLDLESTLQHEHTAALAKREQA 1198
QY 404 QLTAMDOVLGADGELPAELQAIKDALQALAKQ--PSAGGLATAMGOVAFAAKVGGSAA 461
DB 1199 DSYA-----EL-----GEQIDIMQIRVKQKLEKESYKMEIDTLTSMDEVAKAKA----- 1244
QY 462 GTAGTQVMNKKLYTKTAFSTSSSYAALSDGYSAYKT-----LNSLYESRSQV 513
DB 1245 -----NLEKKCKRT-----LEDLQSLSEIKTKSDENVRQLNDM-----MAQ 1277

```

```

QY 514 SAISQGANPALSRVS-----RSGIESQGRSADASQARAETI 550
DB 1278 RARLOTENGESRQLEKEKALVSQLTGRQAOYQOYIEELKRIHEEVRKKNALAHAVQSA 1337
QY 551 VRDSQTLGDVYSRLQVDSIMSTIVS--NPQANQ-----EELMQLTMS 592
DB 1338 RHDCCDLLEQYEEQEKKALQGRMSKANSEVAKWRTKYTEDDAIQRTLEELEKKKLAQR 1397
QY 593 ISKAPQGYPAVONSADSLQKFAQLERE-----FVNGER--SLAESQENAFKRPAPF 643
DB 1398 LQDAEE-SIEAVNSKCSALEKTKORLQGEVEDLMIDIVERANSILA---ANLDKQRNF 1450

RESULT 31
MYSS_RABIT
ID MYSS_RABIT STANDARD; PRT; 1084 AA.
AC P02562;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Myosin heavy chain, skeletal muscle (Fragments).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE OF 1-258.
RA Capony J.-P., Elzinga M.;
RT "The amino acid sequence of A 34,000 dalton fragment from S-2 of
RT myosin."
RL Biophys. J. 33:148A-148A(1981).
RN [2]
RP SEQUENCE OF 259-428.
RX MEDLINE=85131142; PubMed=3972832;
RA Lu R.C., Wong A.;
RT "The amino acid sequence and stability predictions of the hinge
RT region in myosin subfragment 2."
RL J. Biol. Chem. 260:3456-3461(1985).
RN [3]
RP SEQUENCE OF 409-1084 FROM N.A.
RX MEDLINE=87304245; PubMed=3305014;
RA Maeda K., Sczakiel G., Wittinghofer A.;
RT "Characterization of cDNA coding for the complete light meromyosin
RT portion of a rabbit fast skeletal muscle myosin heavy chain."
RL Eur. J. Biochem. 167:97-102(1987).
CC 1- FUNCTION: MUSCLE CONTRACTION.
CC 1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC 1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC 1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC 1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC 1- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
CC CONSERVED.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X05958; CAA29391.1; -
DR PIR; A02985; A02985.
DR PIR; A05280; A05280.
DR PIR; S00084; S00084.

```

KM Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KM ATP-binding; Multigene family.
 FT NON_TER 1 1
 FT DOMAIN <1 >258 ALPHA-HELICAL TAIPIECE (S2).
 FT NON_CONS 258 259
 FT DOMAIN <259 1084 RODLIKE TAIL (S2 AND LMM DOMAINS).
 FT DOMAIN 455 1084 COILED COIL (POTENTIAL).
 FT VARIANT 405 405 L -> V.
 FT VARIANT 408 408 V -> L.
 FT VARIANT 421 421 E -> D.
 FT VARIANT 423 423 S -> G.
 FT VARIANT 426 426 K -> R.
 SQ SEQUENCE 1084 AA; 12548 MW; 229CFD69A6E1F7F0 CRC64;
 Query Match 5.8%; Score 187; DB 1; Length 1084;
 Best Local Similarity 21.4%; Pred. No. 0.69; Indels 174; Gaps 26;
 Matches 145; Conservative 94; Mismatches 265; Indels 174; Gaps 26;

QY 29 QIANSSEKSTKASESPASSSVSSWFLSSAKNALISLSDAILKNSSPTSLSS--- 85
 DB 7 EVANNKEEEKTESLAKAEAKRE-----LEEKWALMOEKNDLOLOVQAEADSLADAE 61
 QY 86 -----OLEASTSTSTVTRAAKDYDEAKSNFDIAKSGLENAKTIAEYETKADIM 135
 DB 62 ERODLIKTKIOLEAKIKEYT-----ERADEDEEINAEILTAKKRKLEDECSLKKDIDDEL 118
 QY 136 AALQDMERLANSDPSNNHVEEVNNI-----KKAL-ENQKOTIDKLNTLVTONQ 184
 DB 119 LAKVEKEKATENKKNLLEEMAGLDETIAKLTKEKKAQOEHAQQLDLDLQ-----AEED 173
 QY 185 KSLTEFLKTTDSADQIPAINSOLEIKRNSADQIINKLEHONISYEAVLNNAGEVIKASE 244
 DB 174 KNTLTAKTKLEQVYDDLEGSLEDEK-----KIRMDLER-----AKRKIE 214
 QY 245 AGIKGQALQSTVDAGDSQAALVLAQONNSPDNIAATKELIDAET---KVNELKOEHT 301
 DB 215 GDLKLAQ--ETSMDEIENDQQLDEKILKLEFMTNLOSKLIEDQALMTNQLTEIELEEEIE 272
 QY 302 GLTDPVIAKKAEEOISQAKDIOETKPSGSDPIYVGPSSASAGSAALAKSSNNSGR 361
 DB 273 AERASR--AKAEKORSUSLELEET-----SRLEBAGA-----TSAQI 310
 QY 362 SLILDDVDNEMAIALQGRSMIEQFNVNNPATAKELQAMEAQLTMSDQLVGADGELPA 421
 DB 311 EM-----NKKREAEKRRDLEENTLOHEATPAALAKKHDSVA--EL--GEQID 357
 QY 422 ETQAIKDALAQALKQ--PSADGLATAMGOVAPRAAKVGGAGTACTYOMNKKOLYKTA 479
 DB 358 NLQRYKOKLEKEKSELMKIDDLA-----GNMETYSKAKGNLEKMCRT-- 400
 QY 480 SSTSSSYAALSDGSAVKTLNLSYSE--SRSGVSAISQTAANPALSR-----S 527
 DB 401 -----LEDOLSEKTKEEHQRLINELSAOKAKLHTESGSEFSQOLDEKDMVSO 449
 QY 528 VSRSG-----TESQGRSADASQRAAETTVRSQTLGDVYSRLQVLDLSMST 574
 DB 450 LSRGQAFQTOIEGLKROLEEFETKAKSALAHALQSSRRCDLLREYEEQEKAKELQ 509
 QY 575 VS--NPQANO-----EETMOKITASISKAPQGYPAVONASDLSQKFAA 616
 DB 510 MSKANSSEVQWMTKCTETDIAQRTTELELEAKKRLAQRLDAEERH-VEAVNSKASLEKTK 568
 QY 617 QLERE---FVDEGERSLA 630
 DB 569 RLQNEAREDMIDVERSNA 586

RESULT 32
 MYHT_PIG STANDARD; PRT; 1935 AA.
 AC P/9293;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
 GN MYH7.
 OS Sus scrofa (pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_Taxid=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Domestica;
 RA Ko T.L.;
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U75316; AAB37320.1; -
 CC DR HSSP: P08799; IAMD.
 CC DR InterPro: IPR000048; IQ_region.
 CC DR InterPro: IPR004009; Myosin_N.
 CC DR InterPro: IPR002928; Myosin_tail.
 CC DR InterPro: IPR001609; myosin_head.
 CC DR Pfam: PF00612; IQ_2.
 CC DR Pfam: PF00612; myosin_head; 1.
 CC DR Pfam: PF01576; Myosin_tail; 1.
 CC DR Pfam: PF02736; Myosin_N; 1.
 CC DR PRINTS: PR00193; MYOSINHEAVY.
 CC DR ProDom: PD000355; myosin_head; 1.
 CC DR SMART: SM00015; IQ_1.
 CC DR SMART: SM00242; MYSC; 1.
 CC DR PROSITE: PS50096; IQ_1.
 CC DR PROSITE: PS50096; IQ_1.
 CC DR Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
 CC ATP-binding; Methylation; Alkylation; Multigene family;
 CC Calmodulin-binding.
 CC KW Calmodulin-binding.
 CC FT DOMAIN 1 780 MYOSIN HEAD-LIKE.
 CC FT DOMAIN 781 810 IQ.
 CC FT DOMAIN 811 810 COILED COIL (POTENTIAL).
 CC FT NP_BIND 839 1935 ATP (POTENTIAL).
 CC FT NP_BIND 178 185 ACTIN-BINDING.
 CC FT DOMAIN 655 677 ACTIN-BINDING.
 CC FT DOMAIN 757 771 ACTIN-BINDING.
 CC FT MOD_RES 129 129 METHYLATION (THI-1) (POTENTIAL).
 CC FT MOD_RES 129 129 ALKYLATION (SH-1) (POTENTIAL).
 CC FT MOD_RES 695 695 ALKYLATION (SH-2) (POTENTIAL).
 CC FT MOD_RES 705 705 ALKYLATION (SH-2) (POTENTIAL).
 CC SQ SEQUENCE 1935 AA; 223110 MW; 149CDBFD910DBB08 CRC64;
 Query Match 5.8%; Score 187; DB 1; Length 1935;
 Best Local Similarity 19.4%; Pred. No. 1.4; Indels 128; Gaps 24;
 Matches 137; Conservative 134; Mismatches 307; Indels 128; Gaps 24;

QY 8 MESGESVSSNQSNMPLINQIANSSEKSTKASESPASSSVSSWFLSSAKNALI 67

```

Db 891 VOAEODNLSDESRCDQLKNKIOLEAKYKEMTERLEDEENNAELTA----- 938
QY 68 SLRDALINKNSPTDLSOLEASTSTSYTRVAAKDYDEAKSNFPAKSGLEMANLTALEY 127
Db 939 -----KKRVEDEBCESEKLRDIDDELTL--LAKVEKKHAKTENKVNKLTEEMAGLDEI 988
QY 128 ETKMADLMAALODMERLANSDPNSNHTEEVNNIKKALEAKQDITDKL----- 174
Db 989 IAKLTKREKALQEAHQQA--LDLOLQAEEDKVNLTAKAKVLEQHVDDLAGSLBOEKKVRMD 1047
QY 175 -----NKLVTLQONKSLTEVLKTTDSADQIPAINSOLEIKNSADQII 218
Db 1048 LERAKRKLEGLDKLQESTIMDENDKQQLDERLKKKF--ELNALNARLEDQALCSQQL 1105
QY 219 KDL-----ERONISYEAVLTNAGEVIRKASSEAGIKLQALOSIYDAGDOSQAALVLAQ 271
Db 1106 KKLKELQALIELELELELEKERTAKAVEKLRSDLSRELEISERLEEAAGATSVQI--EMN 1164
QY 272 QNNSPDNIATKELIDAA--ETKYNELKQEHGTGLDPLVKAEBOISQAOKDIOEIKP 328
Db 1165 KKRBAEFQKMRDLBEATLQHEATAALRRKH--ADS--VALGRQIDNLQRVKQKLEK 1219
QY 329 SCSDIPIVGPSSASAGSAGACALSSNNSGRISLLDQVNDMAIALQGRSMIE--- 385
Db 1220 EKSEFKELDDYTS---NMEOIRKAKANLEKMCRTLEQDMEHRSKAEETORSVNDLTS 1275
QY 386 ---QFVNVNPAATKELQAMEA---OLT---AMSDQVYAGDELPRPEIAIDALQALK 435
Db 1276 QAKAKIQTENGELSRQDEDEALISQLTRKGLTYTQOLEDKRLQEEVKA--KNALAHALQ 1334
QY 436 --QPSADGLATMGQVAFAAKVGSG--SAGTAGVOMNVK---OLYKTAFSSTSSSYA 488
Db 1335 SARHADLRLQEQEETETKAELOVRVLSKANSEVAQWRKYETDAIQRTEELEAKKILA 1394
QY 489 AALSDGYSAVKTLN---SLYESRSGVQSAI-----SOTANPALSRSVS----- 529
Db 1395 ORLODAEEVEAVNANAKSSLEKTKHRLQNEIEDLVNVERSNAAAALDKORNFDKILA 1454
QY 530 --RSGLIESGRSADSORAETIVRDSQTLGQVSRLOYLDSIMSTIVSNPQANOBEIMO 587
Db 1455 EMKQKKEESQSELESQKARSLSTELFKLKNYE--ESLEHLETKRREN--KNLOBEIS 1510
QY 588 KLTASISKAPQFGYPAVONSADSLQFAAQLEREFEVNGERSIAESQ 633
Db 1511 DLTEQGGSGK-----TIHELEKVRKQLEAKLEQLSALDESE 1548

RESULT 33
MFP1_LYCES
ID MFP1_LYCES STANDARD: PRT: 697 AA.
AC P93203:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MAR binding filament-like protein 1.
GN MFP1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv VENT Cherry; TISSUE=fruit;
RX MEDLINE=97112038; PubMed=8953774;
RA Meier I., Phelan T., Grunsem W., Spiker S., Schneider D.;
RT "MFP1, a novel plant filament-like protein with affinity for matrix
RT attachment region DNA.";
RL Plant Cell 8:2105-2115(1996).
CC -I- FUNCTION: BINDS DNA. INTERACTS WITH CHROMATIN VIA MATRIX
CC ATTACHMENT REGIONS (MARS). LIKELY TO PARTICIPATE IN NUCLEAR
CC ARCHITECTURE BY CONNECTING CHROMATIN WITH THE NUCLEAR MATRIX AND

```

```

CC CC POTENTIALLY WITH THE NUCLEAR ENVELOPE.
CC -I- SUBCELLULAR LOCATION: NUCLEAR MATRIX.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL: Y07861; CAB69181.1;
KW Nuclear protein; DNA-binding; Coiled coil.
FT DOMAIN 125 681 COILED COIL (POTENTIAL).
FT DOMAIN 10 15 POLY-SER.
SQ SEQUENCE 697 AA; 79516 MW; 700A56D68D6A7E49 CRC64;

Query Match 5.8%; Score 186; DB 1; Length 697;
Best Local Similarity 19.7%; Pred. No. 0.44;
Matches 98; Conservative 84; Mismatches 151; Indels 164; Gaps 19;

QY 29 QIASNETKESRKASBPASSSVSWGFLSSAKNALISL-----RDALINKNS---SPT 81
Db 246 QVLEEKITLITTEIKDKVEYSLNSNTSKLAKESEVNSLSDMYQOSODOLMNLSEIKELX 305
QY 82 DLSOLEASTSTSYVT-----RVAKDYDEAKSNFDTAKSGLE 119
Db 306 DETOKRERELKCVSEDNLVNQLNSLLERDESKELHAIQKEYSEFKNSND---EKVA 362
QY 120 NAKTALAEYETKADLMAALODMERLANSDPNSNH-----TEEVNNIKKALEAQDITDK 173
Db 363 SATTGEQEKRLHQ---LEBQGLTALSEASKNEVLIADLTREKELRMRVALEIDNVNK 418
QY 174 LNK--LVTLQONKSLTEVLKTTDSADQIPAINSOLEIKNSADQIIKLERONISYE-- 229
Db 419 LKQELFVTEBSELSNRSKSEVSDITVQLELRDLSKIE---REVSKLQMLEETRSLQGN 475
QY 230 -----AVLTNAGEVIRKASSEAGIKGLQALQISTVDAGDOSQAALVLAQOQNSPD 277
Db 476 IDETKHSSELALAECLTYYTKELLKTNE-----EMHTMSD 509
QY 278 NIATKELIDAAETKY-----NELKQEHGTLSPIYKKAEOISQAOKDI--- 323
Db 510 ELVAAYSENDSLOTLELVNYYKKRREHTRNELKQEK---IVRTLEELKLELSQITRE 563
QY 324 QETKPSGSDIPIVGPSSASAGSAGACALSSNNSGRISLLDQVNDMAIAL----- 376
Db 564 KELRKSLED-----ELEKATES-----LDEINRNVLALELELAT 599
QY 377 ----LQFR---SMIDQFVNPNPA-----TAKEIQAMEAQLTMSDOLV 413
Db 600 SRNSLEDEREVHRQSVBOKISOEAQENLEDAHSLVNLKLERESLEKRAKKLEDEMA 659
QY 414 GADGE---LPAELQAIK 427
Db 660 AAKGELLRLRSQINSYK 676

RESULT 34
MYHB_HUMAN
ID MYHB_HUMAN STANDARD: PRT: 1972 AA.
AC P35749; O00396; P78422; O94944;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMMC).
GN MYH11 OR KIA0866.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=99425270; PubMed=10493829;
 RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
 RA Furlmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
 RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
 RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
 RT "Genome duplications and other features in 12 Mb of DNA sequence from
 RT human chromosome 16p and 16q.";
 RL Genomics 60:295-308(1999).
 RN [2]
 RP SEQUENCE OF 1-1266 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:355-364(1998).
 RN [3]
 RP SEQUENCE OF 885-1972 FROM N.A.
 RC MEDLINE=99263189; PubMed=7684189;
 RX Matsuo K., Yoshida M.C., Fututani Y., Imamura S., Kanda N.,
 RA Yanagisawa M., Masaki T., Takao A.;
 RT "Human smooth muscle myosin heavy chain gene mapped to chromosome 1
 RT region 16q12.";
 RL Am. J. Med. Genet. 46:61-67(1993).
 RN [4]
 RP SEQUENCE OF 1093-1972 FROM N.A.
 RC TISSUE=Hippocampus;
 RX Okaïma K.;
 RA Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
 RL
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SURCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE; EXPRESSED IN THE UMBILICAL
 CC ARTERY, BLADDER, ESOPHAGUS AND TRACHEA.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC AKTYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- DISEASE: A chromosomal rearrangement, known as pericentric
 CC inversion inv(16)(p13q22), produces a fusion protein that consists
 CC of the 165 N-terminal residues of CBF-beta (PPB2) with the tail
 CC region of MYH11. This rearrangement is associated with acute
 CC myeloid leukemia of M4EO subtype.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MYRMOYOSIN (LMW) AND 1 HEAVY MYRMOYOSIN (HMW). IF CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF001548; AAC31655.1; -
 DR EMBL: U91323; AAC35212.1; -
 DR EMBL: AB020673; BAA74889.1; -
 DR EMBL: D10667; -; NOT_ANNOTATED_CDS.
 DR EMBL: X69292; CAA49154.1; -
 DR HSSP: P10587; 1BR2.
 DR Genew: HGNC:7569; MYH11.
 DR MIM: 160745; -
 DR InterPro: IPR000048; IQ_region.
 DR InterPro: IPR004009; Myosin_N.

```

QY 550 IYRDSOT----LGDYISRLQVLDLSMTSTVSNPQNOEIMOKLTASISKAPQGYPAVQ 605
DB 1362 LERHISTLNIOISDSKKKLDPASTVEALEGKKRQKEI-ENLTQO-----YEKA 1412
QY 606 NSADSLQKFAAOLEREF-----VDGERSLASEQENAFKQPAFIQOVLVNIASLPSGY 658
DB 1413 AAYDKLEKTKRNLOQLDLDVLDNQNOLVSNLEKKOKR----FQDLAEKEKNISKY 1467

RESULT 35
AKA9_RABIT
ID AKA9_RABIT STANDARD: PRT: 1087 AA.
AC Q28628:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
DE (PRA9) (A-kinase anchor protein 120 kDa) (AKAP 120) (Fragment).
GN AKAP9 OR AKAP120.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Gastric parietal cell;
RX MEDLINE=97220389; PubMed=9148752;
RA Dransfield D.T., Yeh J.L., Bradford A.J., Goldenting J.R.;
RT "Identification and characterization of a novel A-kinase-anchoring
RT protein (AKAP120) from rabbit gastric parietal cells.";
RL Biochem. J. 322:801-808(1997).
CC -I- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE
CC A. MAY BE A SCAFFOLDING PROTEIN.
CC -I- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN GASTRIC PARIETAL CELLS.
CC -I- DOMAIN: RII BINDING SITE. PREDICTED TO FORM AN AMPHIPATHIC HELIX,
CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U26360; AAC35413.1; ALT_INIT.
CC
CC Coiled coil.
CC KM NON_TER 1 1
CC FT DOMAIN 559 572 PRA-RII SUBUNIT BINDING DOMAIN.
CC FT DOMAIN 5 461 COILED COIL (POTENTIAL).
CC FT DOMAIN 614 773 COILED COIL (POTENTIAL).
CC FT NON_TER 1087 1087
CC SQ SEQUENCE 1087 AA; 124756 MW; 9D916BEOCA9FP02 CRC64;

Query Match 5.8%; Score 185.5; DB 1; Length 1087;
Best Local Similarity 22.3%; Pred. No. 0.8;
Matches 139; Conservative 98; Mismatches 242; Indels 145; Gaps 27;

```

```

DB 335 EEKNELIRPLEQIECL-----MSDQERYKNREBEITQNLNEVIEKLOELANIDOKTS 388
QY 336 VGPSSASAGSAGALKSSNNSGRISLLDDVDNEMAAIALQGRSMIEQFNVPNPATA 395
DB 389 VDPS-----SLSEADSLKHQ-----LDKVIKELALEHQ-----VETNEEMAVTK 430
QY 396 KELQAMEAQLTMSDL- VGDGELPAETQAI-----KDA 429
DB 431 NVLKETNFKNMOLTOELSLKREEREMERIOGPEKSVNMSVGLSKDKPENDLIPTEDA 490
QY 430 LAQALROP-----SADGLATAMGOVAPAAKVGSGAGTAGTVOMVVKOLYK-- 476
DB 491 LQOLETQTOLRSSSESSSKVSLSLKTLQL-----ESTVSTDELITQCYKOI 539
QY 477 -----TAFSSSTSSSSVAALSDGYSAYKTINSLYSESR-SGVQSAISQTN-PAL 524
DB 540 QDMREGRETEMLQTKIYSLQVLEE-----KVAALVSQVLEAVQGYVKLADKPAY 594
QY 525 SRSVSRSGIESQGRSADASQRAETIVRDSQTLGDIYSLQ-VLDSLMTIYSNQANDE 583
DB 595 SSDPATEVPG-----LSQLAGNTEMSPVSALTWRISELQSQVEMHSSLIS--EKQV 646
QY 584 EIMOKLTASISKAPQGYPAVONSADSLQKFAOIERE---FVDGPRSLAES----- 632
DB 647 EIAEKNALEKEKKLQELQKLVQDSE-----TKQREKRSKRLHGDGLVLESTSESGV 700
QY 633 --QENAFRRQPAFIQOVLVNIASL 654
DB 701 FQELALRASAAPKGLANVEL 724

RESULT 36
SMC_MYCTU
ID SMC_MYCTU STANDARD: PRT: 1205 AA.
AC Q10970:
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromosome partition protein smc.
GN SMC OR RV2922C OR MT2990 OR MTCY338.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Cobbe N., Heck M.M.S.;
RT "Phylogenetic analysis of SMC proteins.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Bartell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;

```


KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
 FT DOMAIN 78 120 CAP-Gly 1.
 FT DOMAIN 143 204 SER-RICH.
 FT DOMAIN 232 274 CAP-Gly 2.
 FT DOMAIN 304 331 SER-RICH.
 FT DOMAIN 350 342 COILED COIL (POTENTIAL).
 FT DOMAIN 1408 1421 CHC-BOX.
 FT VARSPLIC 457 491 MISSING (IN SHORT ISOPFORM).
 FT CONFLICT 1069 1069 D -> E (IN REF. 2).
 SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;
 Query Match 5.8%; Score 185.5; DB 1; Length 1427;
 Best Local Similarity 20.5%; Pred. No. 1.1;
 Matches 143; Conservative 116; Mismatches 291; Indels 147; Gaps 28;
 QY 5 HHNMGSPSPVSSNOSSMPITNGOIASNSEPKESFKASEAS--PSASSSSMSRLSSA 62
 DB 276 HAYTKIGFSTTPKAKANAVRVNATTSALKRSPSASSLSMSVASSVS---RPS 331
 QY 63 KNAISLRDALINKNSPTDSLQLEASTSTSTVTRVAAKDDEAKSNFDYAKSG-LENA 121
 DB 332 RIGLLETSSRVARKISGTTALQELKEKQOHTEQLLAERDLERAVAKATSHVGEIEQ 391
 QY 122 KTLA-----EYTKADLMALQDMERLANSPPSNHTPEVNNIKKALEAQDITD 172
 DB 392 LALARDGHQVHLELAKMDQLTWTEADR-----EKVELLNQ-LEEEKRYVE 439
 QY 173 KINKLVTLLONONKSLTEVTKTSDAQIPAINSLINKNSDAQIILKDERQISAEVY 232
 DB 440 DLQFRKEESTITKDELETQTKLEHA-RIKELEBSLFEKTKADKIDRELEDTIV---AIV 495
 QY 233 TNAGEVIRKASSENGIK-----LGAQLQSTIVDAGD-----QSOAAVLQAOQNN 275
 DB 496 SEKSRIMELEKDLALRQVEAELRRLRLESNKRPAGDVMSLSLQETISSIQEKLVTRTD 555
 QY 276 PNIIATKELIDAE-----TKVNE---LKOEHGLDPSIV---KK 311
 DB 556 QREITSLKEHFGARETHOKERKALYATATEKLSKENSELSKLEHANKNSDVIALMKSK 615
 QY 312 AEEQISOAQKDIQIEKPSGSDIPYVPSGSASAGSAGAKSSNNSGRISLLDDVDY 371
 DB 616 LERAIASHQOAMELKVS-----KLGITTAFA-ELKQIEKML-----DYQHE 662
 QY 372 MAALALQFRSMTEQFNVNVPATAKELQAMEAQLTAMSDQLVAGDELPAEIQAIKDATA 431
 DB 663 I-----ENLNOODSEERAAHAKEMEARAKIMKYIK-----EKENSLAIRSKLD 707
 QY 432 QALKOPSDGLATAMGOVAFPAAKVGGSGAGTAGTYOMNKOLYKTAFFSTSS----- 484
 DB 708 KAEQDHLVE-MEDTLNKLQDAETKV-----KELEVLQAKCNEOTKYIDNFTSOLKATEEK 761
 QY 485 ----SSYAAALSDGYSAVYKTLNLSYSESRGVOS-AISQTANPALRSRVRSIESQGRS 539
 DB 762 LDDLALKRASSRSGKEMMKLRQOLEAKEQIKHLEIKNAESSKASSITR---ELQGRE 818
 QY 540 -----ADASQRAAETIVDSQTLGDVSR-----LQVLDLSMTSTVSNPQANQEE 564
 DB 819 LKLTNLQENISEVSO-VKETLEKELQILKEKFAEASEEAVSVRSQGETV--NKLHQKEE 875
 QY 585 IMOKLTASTISKAPQPGYPAVQNSADSLQKFAADLERE 621
 DB 876 QFNMLSSDLEKLR-----ENLADMEAKFRERDERE 905
 RESULT 38
 MYSC_CAEEL
 ID MYSC_CAEEL STANDARD; PRT; 1947 AA.
 AC P12845;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain C (MRC C).
 GN MYO-2.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=89178677; PubMed=2926820;
 RA Dibb N.J., Maruyama I.N., Krause M., Karn J.;
 RT "Sequence analysis of the complete *Caenorhabditis elegans* myosin
 heavy chain gene family";
 RL J. Mol. Biol. 205:603-613(1989).
 RN [2]
 RP SEQUENCE OF 1-21; 112-371 AND 1501-1772 FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=85201409; PubMed=3888374;
 RA Karn J., Dibb N.J., Miller D.M.;
 RT "Cloning nematode myosin genes";
 RL Cell Muscle Motil. 6:185-237(1985).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
 C. ELEGANS.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: X08066; CAA30855.1; -
 DR EMBL: M37233; AAA28121.1; -
 DR EMBL: M37235; AAA28122.1; -
 DR EMBL: M37236; AAA28123.1; -
 DR PIR: S05697; S05697.
 DR HSSP: P10587; 1BR2.
 DR InterPro: IPR004009; Myosin.N.
 DR InterPro: IPR002928; Myosin.tail.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF01576; Myosin_tail; 1.
 DR Pfam: PF02736; Myosin_N; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.
 DR SMART: SM00242; MSc; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family.
 FT DOMAIN 1 855 MYOSIN HEAD-LIKE.
 FT DOMAIN 856 1947 RODLIKE TAIL (S2 AND LM DOMAINS).
 FT NP_BIND 174 181 ATP (BY SIMILARITY).
 FT DOMAIN 668 690 ACTIN-BINDING.
 FT DOMAIN 773 787 ACTIN-BINDING.
 FT MOD_RES 125 125 METHYLATION (TR1-) (POTENTIAL).
 FT MOD_RES 708 708 ALKYLATION (SH-1).
 FT MOD_RES 718 718 ALKYLATION (SH-2).
 FT CONFLICT 132 132 E -> D (IN REF. 2).

Db 560 TTTLATANNSTRAATVITDPTLPTDNDASPD-----NAKSTYSGSSSTGASL 611
 QY 346 GS---AAGALKSSNNSGRISLLDDVD-NEMAAIALOGFRSMIEQFVNVPATAKELQAM 401
 Db 612 DLRTTTSISVSSNMTQVSTCTSESDYSDSPFAI-----STATTEENLI 658
 QY 402 EAOLITMSDVLGADGLPAEIOAIKALA-----QALKPSADGLATAMQOVAFPAAKV- 456
 Db 659 TMTTAS---CSTDSNFPSPSAASSTDETAFTRTISTSCSLNGASQTSLELTSPKTN 714
 QY 457 -----GGGSAGTACTVQNNVHQLYKTAFFS-----TSSSYAALSDGYS 497
 Db 715 TVVPASSFPSTTTTCLEN-----DDTAFSSSYTEVNATITINPGTSSSLADDFATSEKPN 770
 QY 498 YKTLNLSYSSRSRGVOSAISQT-----ANDALSRVSRSGIESQGRSADASQRAAETIVR 552
 Db 771 PTSVKSTSNEGSTSTTTTYQVATVIAKPSSTSLGARITGTSNGRSTTSQDDCSAM--- 827
 QY 553 DSGTLGIVSRLO 565
 Db 828 -HOPTSISYVQLK 839

RESULT 40
 MYH4_HUMAN
 ID MYH4_HUMAN STANDARD; PRT; 1939 AA.
 AC Q9Y623;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, skeletal muscle, fetal (Myosin heavy chain 11b) (MyHC-11b).
 GN MYH4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=99318869; PubMed=10388558;
 RA Weiss A., Schiaffino S., Lehmann L.A.;
 RT Comparative sequence analysis of the complete human sarcomeric myosin heavy chain family: implications for functional diversity.;
 RL J. Mol. Biol. 290:61-75(1999).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TO DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: AF111783; AAD29949.1;
 DR HSP; P13538; 2MYS.
 DR GeneW; HGNC:7574; MYH4.

DR MIM; 160742; IO_region.
 DR InterPro; IPR000048; Myosin_N.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family.
 FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
 FT DOMAIN 785 814 IQ.
 FT DOMAIN 843 1939 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP (POTENTIAL).
 SQ SEQUENCE 1939 AA; 223012 MW; 40B1AD1D77A47DE CRC64;

Query Match 5.7%; Score 184; DB 1; Length 1939;
 Best Local Similarity 19.9%; Pred. No. 1.9; Indels 122; Gaps 22;
 Matches 113; Conservative 104; Mismatches

QY 102 KDVEAKSNFDTAKSGLENKTLA---EYETKMAIDLMAALQDMERLANSDDPSNNHTIEVN 158
 Db 851 KEANAKKEFEKTRKE-ELAKTEAKKRELEKAVTLM-----QENK 889
 QY 159 NIKKALEAQKDTI---DKLNKLVTLQNNKSLTEVLKTTDSADQIPAINSOLEINKNSA 214
 Db 890 DLQVQAEADALDAEERCDQLIKTKIQLEA--KIKEVTERADEBEINAEITAKRKRL 947
 QY 215 DQIIKDLERNITSEAVLTVAGEVIRKASSPAGIK-----LGQALQSVVDAGDQSQAA 266
 Db 948 EDECSLEKIDIDLETLTAV-EKEKHATENKVKNLTEENAGLDETAKLT---REKKA 1002
 QY 267 VLOAQNNNSDNIATATKELIDAEETKYNELKQEHGTGSDPLVKKAEQISQAQKDIQEI 326
 Db 1003 LQEAHQTLDD-----LQMEDKAVNTLTAKT-----KLEQVDDLEGSLEOC 1045
 QY 327 KPSGSDPIVPGSASASAGSAGALKSSNNSGRISILLDDVNEMAAIALOGFRSMIEQ 386
 Db 1046 KKLQMDL-----ERAKRKLGGDLKLAQESTM-----DFENDKQQLNEKLRKEFEW 1091
 QY 387 FVNANPATAKELQAMEQLTAMSDQVAGDELPAEIOAIKDALQALQPSADGLATAM 446
 Db 1092 SNLQ--GKIDEDALAMQLOKKIKELQARIELEEBIEAERASRAKAKORS--DLSREL 1147
 QY 447 GOVAFAAKAVGGAGTACTVQNNVHQLYKTAFFSSTSSSYAALSDGYSAVYTLNLSYS 506
 Db 1148 EELISERLEEGG---ATSAQIELNKR-----EAEFQ 1176
 QY 507 ESRSGV-QSAISQATNPALSRVSRSGIESQGRSADASQRAAETIVRDSQTLGIVSRLO 565
 Db 1177 KMRDLEESTLQHEATPAALRKKHADSVAGELQIDSLQVKKLEKESKL-----KME 1231
 QY 566 VLD--SLMSTIVSPQANQEIIMOKLTASISKAPQGYPAVQNSADSLQFAQLEREFV 623
 Db 1232 INDLASNMET-VSKAKNPFKMCRTLEDQSLI-KTKEERQRLINELSAQKARLHTESG 1289
 QY 624 DGERSLAESQ---ENAFRKQPAFIQOV 647
 Db 1290 EFSRQDEKDAWYQSLSRGKQAFITQOI 1316

Search completed: March 12, 2003, 12:10:27
 Job time : 74 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 11:23:47 ; Search time 40 Seconds

(without alignments) 1566.218 Million cell updates/sec

Title: US-10-007-693-139

Perfect score: 3204

Sequence: 1 MHHHHHHMESGESVSSNQ.....PAFIQVLVNIAISLFGSL 660

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2943	91.9	647	2	hypothetical prote
2	894	27.9	651	2	conserved hypothet
3	894	27.9	651	2	CHLPN 76 kDa homol
4	453	14.1	715	2	hypothetical 76k p
5	268	8.4	2271	2	hypothetical prote
6	261	8.1	6713	2	hypothetical prote
7	235	7.3	2285	2	probable transglyc
8	234	7.3	1063	2	hypothetical prote
9	231	7.2	1365	2	hypothetical prote
10	226.5	7.1	1269	2	limp1 protein - Myc
11	225.5	7.0	1302	1	probable myosin he
12	224.5	7.0	1790	2	surface-located me
13	219.5	6.9	2139	2	transport protein
14	218.5	6.8	2570	2	myosin heavy chain
15	218	6.8	4776	2	fibritiae-associate
16	216.5	6.8	2055	2	cell wall surface
17	214	6.7	1072	2	extracellular matr
18	213	6.6	2186	2	hypothetical prote
19	212	6.6	1690	2	hypothetical prote
20	211.5	6.6	1822	2	microtubule bindin
21	210.5	6.6	2481	2	EF protein - Strep
22	210	6.6	1091	2	FmbB protein (limp
23	208.5	6.5	1104	2	probable membrane
24	208	6.5	1133	2	hypothetical prote
25	208	6.5	1306	2	probable membrane
26	208	6.5	1475	2	MSB2 protein - yea
27	208	6.5	3890	2	hypothetical prote
28	207.5	6.5	641	2	hypothetical prote
29	205.5	6.4	2007	1	methy-accepting c
					myosin heavy chain

30	204	6.4	1051	2	T18351	limp1 protein - Myc
31	202.5	6.3	1963	1	MMKW	myosin heavy chain
32	202	6.3	1940	2	A59287	myosin heavy chain
33	201	6.3	2116	2	A28635	myosin heavy chain
34	200	6.2	971	2	B90835	probable tail fibre
35	200	6.2	973	2	C85693	probable membrane
36	200	6.2	1992	1	S02771	myosin heavy chain
37	200	6.2	2052	2	C97038	phage-related prot
38	199.5	6.2	1381	1	S45781	probable calcium-b
39	199.5	6.2	1957	2	T38077	hypothetical colle
40	199.5	6.2	2253	2	T30336	nuclear/mitotic ap
41	199	6.2	918	2	C96829	unknown protein F1
42	199	6.2	1509	1	A27224	myosin heavy chain
43	199	6.2	1940	1	A24922	myosin heavy chain
44	198	6.2	1127	2	T28317	ORF MSY156 hypothe
45	197.5	6.2	890	2	S22452	surface exclusion

ALIGNMENTS

RESULT 1									
G71490									
hypothetical protein CT622 - Chlamydia trachomatis (serotype D, strain UW3/Cx)									
N:Alternate names: chlpn 76kda homolog CT622									
C:Species: Chlamydia trachomatis									
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999									
C:Accession: G71490									
R:Stephens, R.S.; Kaiman, S.; Iammell, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche									
Science 282, 754-759, 1998									
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t									
A:Reference number: A71570; MUID:9900809; PMID:9784136									
A:Accession: G71490									
A:Molecule type: DNA									
A:Residues: 1-647 <ARN>									
A:Cross-references: GB:AE001333; GB:AE001273; NID:q3329068; PIDN:AAC68226.1; PID:q332									
A:Experimental source: serotype D, strain UW-3/Cx									
C:Comment: This sequence was originally identified as homologous to part of a sequenc									
PIR:E72042).									
C:Genetics:									
A:Gene: CT622									
Query Match									
Best Local Similarity 91.9%; Score 2943; DB 2; Length 647;									
Matches 613; Conservative 15; Mismatches 19; Indels 6; Gaps 1;									
QY	8	MESGESVSSNQSNMPTIINGQIASNETKESTKASEASPSASSVSSWFLSSAKNALI	67						
DB	1	MESGESVSSNQSNMPTIINGQIASNETKESTKASEASPSASSVSSWFLSSAKHALI	60						
QY	68	SLRDAILNKNSPTDLSGLEASTSTSTVTRAAADYDAKSNFPTAKSGLENATTLAAY	127						
DB	61	SLRDAILNKNSPTDLSGLEASTSTSTVTRAAADYDAKSNFPTAKSGLENATTLAAY	120						
QY	128	ETKMAIDMAALQDMERLANSNHNTEVNIKKALAEQKOTIKLKLVTYLNQNNKSL	187						
DB	121	ETKMAIDMAALQDMERLANSNHNTEVNIKKALAEQKOTIKLKLVTYLNQNNKSL	174						
QY	188	TEVLKTTSDADQIPAINSOLEIKNKSADQIKDLERONISYPAVLTNAGEVIKASSEAGI	247						
DB	175	TEVLKTTSDADQIPAINSOLEIKNKSADQIKDLERONISYPAVLTNAGEVIKASSEAGI	234						
QY	248	KLGQALQSTIVDAGDQSOAAVLAQOONNSPDNIAAKRELIDAEETVYNELKQHTGTLTSP	307						
DB	235	KLGQALQSTIVDAGDQSOAAVLAQOONNSPDNIAAKRELIDAEETVYNELKQHTGTLTSP	294						
QY	308	LYKKAEDIOISQAKDIOETIKPSGSDIPTVPGSAAAGSAGSAGALSSNNSGRISLLDD	367						
DB	295	LYKKAEDIOISQAKDIOETIKPSGSDIPTVPGSAAAGSAGSAGALSSNNSGRISLLDD	354						
QY	368	VDNEAAIALQGRSMIEQFNVNPNATPAKEIQAMEAQITLMSDQLVGADGELPAETQAIK	427						
DB	355	VDNEAAIALQGRSMIEQFNVNPNATPAKEIQAMEAQITLMSDQLVGADGELPAETQAIK	414						

QY	428	DALAAALKQPSADGDIATAMGOVAFPAAAKVGSGSAGTACTVOMNKQLYKTAFSSSTSSSY	487
Db	415	DALAAALKQPSIDGDIATAMGOVAFPAAAKVGSGSAGTACTVOMNKQLYKTAFSSSTSSSY	474
QY	488	AAALSDGSAVKTLLSLTSESRSYGOSAISOTANALRSVSRSGISESGSADASQRAA	547
Db	475	AAALSDGSAVKTLLSLTSESRSYGOSAISOTANALRSVSRSGISESGSADASQRAA	534
QY	548	ETIVDSQTLGDVYSRLVDLSLMTIVSNPOANDEELMOKLTASISKAPOFGYPAYONS	607
Db	535	ETIVDSQTLGDVYSRLVDLSLMTIVSNPOANDEELMOKLTASISKAPOFGYPAYONS	594
QY	608	ADSLQKFAAQLEREFPVDEGERSLAESQENAFRKQPAFIDQVLVNIASLFSGYLS	660
Db	595	ADSLQKFAAQLEREFPVDEGERSLAESQENAFRKQPAFIDQVLVNIASLFSGYLS	647

RESULT 2

D72042

conserved hypothetical protein CP0018 [imported] - Chlamydomophila pneumoniae (strains CMT

N:Alternate names: ch1pn 76 kda homolog_1 (ct6622); hypothetical protein CP0728

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: D72042; D81623

R:Kaltman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Pan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Molecule type: DNA
A:Residues: 1-651 <ARN>
A:Cross-references: GB:AE001654; GB:AE001363; NID:g4377031; PIDN:AAD18867.1; PID:g4377031
A:Experimental source: Strain CML029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gail, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: DB1623
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-651 <REA>
A:Cross-references: GB:AE002165; GB:AE002161; NID:g7188948; PIDN:AA37914.1; PID:g7188948
A:Experimental source: strain AR39, HL cells
C:Comment: This sequence was originally identified as homologous to part of a sequence (PIR:H71490).
C:Genetics:
A:Gene: CPn0728; CP0018

	Query Match	27.9%; Score 894; DB 2; Length 651;
	Best Local Similarity	34.6%; Pred. No. 8.3e-31;
	Matches 233; Conservative 126; Mismatches 262; Indels 52; Gaps 17;	
OY	11 GPESVSNMOSSNPIINQIASNSETKSTKA-----SEAS--SASSSVSMELSSAK 63	
Dd	6 GGPPIDETERTPPADLSAQGLEPASAANKSAEQRIRAGAEKREKSTDSVERWSLTRSAV 65	
OY	64 NALISLRD--ATLNKNSSPTDSLQLEASTSTYTTRAAKDYGDAKNPFPAKSGLDNA 121	
Dd	66 NALMLSLDLKLGIASSNSSSTSRSS--ADVDSSTAATAPTPPPEPDYKKQAQTAFTDTITS 124	
OY	122 KTLALETFRMADIMMALODMERLANSDPSNNHTEEVNNIKALEAOKTIID-----K 173	
Dd	125 TSLADIQAAIVLSLQAVYNINKDTAATD-----EET-----AIAEWETKNADAYKVAQ 173	
OY	174 ENKLVTLONQNSLTLEVAKTTTDSADQIPAINSOLEINKNSADOIIKDLERONI---SYEA 230	
Dd	174 ITELAKYASNDQAIULDSIGKITLSPDLLQALLQSVANNKNNAKEILLKENGDMPNVVGKTPA 233	
OY	231 VLTNNGEVIKASSEGKICGALOSITVDAGDOSCAAVILOAOQNNSPNDIATKELIDAAE 290	
Dd	234 IAQSILVDDTDATATOTIEKDGNALRIDAYFAGONASCAVENAKSNSITSIDSKAALIAAK 293	

QY	291	TKVNEIKOEHCITLTSPLYKKAEEQISOAKCITQELKRS-GSDITIVGPS-GSAASAGSA	348
Db	294	TOIAAROK-PPSPITLIEAEQWVIAEKELKNIKIPKADGSDVNPPTGVGSGXQOGS	350
QY	349	AGALRKSNNSGRISLLDDVDNEMAIALOGFRSMTEOFRNVNPPATKLOAMEKOLTM	408
Db	351	IGST-----RVSMLLDDAENETASILMSGRQMTIHFNTEPNPDSQAQDELAQAARA	403
QY	409	SDOLVGADELPALIEQAIKDALAQL-KOPSADGLATAMGQVAFAAKVGSGSAGTACTV	467
Db	404	K--AAGDDSAALADAQKALEALGKAGQOQGITLMLGQIASAAVVSAGVPAASIS	460
QY	468	QMNVOIKLTARSSSSSSSYAALSDGSAYTILMSLISEMS-GVQASISTANPALSR	526
Db	461	GSSVQILKYT--SKTSGDYKTQISAGDYAKSINDAYGRANNDATRDVINNVSTPLTR	518
QY	527	SYSRSGIESOGRSADASQPAETIVRDSOTLDDVYSRLOVDSLSTIASYSPQANOCEIM	586
Db	519	SVPRARTERAG-PEKTDQALARVIGSNRITLDDVYSQVSAALSVQOIIOSNQANNEIR	577
QY	587	OKTASISKAPQFGYPAYVANSADSLQKFAAOLEREFEVDGERSLASQENAFRKQAPATQO	646
Db	578	OKTASVATKPPQFGYPVQLSNDSTQKFIAKLESFLAEGSRPAAEIKALSFETNSLFIQO	637
QY	647	VLVNIASTLESGYL	659
Db	638	VLVNIIGSLYSGL	650

```

RESULT 3
E86581
CULP7 76 kDa homolog_1 (CT622) [imported] - Chlamydephila pneumoniae (strain J138)
C:Species: Chlamydephila pneumoniae; Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: E86581
R:Shital, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Uuchi, K.; Shiba,
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: AB6491; MUID:20330349; PMID:10871362
A:Accession: E86581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <STO>
A:Cross-references: GI:BA000008; NID:g8979100; PIDN:BA089835.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPJ0728

```

Query Match	27.9%	Score 894	DB 2	Length 651
Best Local Similarity	34.6%	Pred. No. 8.3e-31		
Matches 233	Conservative 126	Pos. Matches 262	Indels 52	Gaps 17
QY	11	GPESVSNSSNNPIINGOIASNSETKESTKA-----SEASP--SASSSVSSMSFSLSSAK	63	
DB	6	GGCPIDETERTPPADISAGLEASAAKAEQRIAGAEAKPRESTDSVERKSIIRSV	65	
QY	64	NALISLRD--AIIUNKNSPTDLSOLEASTSTSTYTRVAAKYDEAKSNFTPAQSLGENA	121	
DB	66	NALMSIADKLGLTASSSSSSSTRS--ADVDTTATATAPTPPPPFDDKYTAQRIAYDTIFTS	124	
QY	122	KTLAEYETKMAIDMALQDMERLANSPPSNHTEEVYNNIKALEAQDFTID-----K	173	
DB	125	TSIADIGALVLSQDAVNTIKDPTAATD-----EET-----AIAAEWETKNDADVKGQ	173	
QY	174	LKKLVATLQONKSLTEVILKTTSDADQIPAINSOLEINKKSADQIIKDLERQNI---SYPA	230	
DB	174	ITELAKVYASDNOAIIIDSLGKLTSPFLLQALALQSVANNKKAELKEKMODNPVEVPEKTPA	233	
QY	231	VLTNAGEVTKASSEAGIKRIGALQSLQVDDGDSQAANVLOAQONNSPDNIATARKELIDAE	290	
DB	234	IASQIVDDQDAATQATELEKQGNARIRAFYVGGQNASGVENAKSNNSISNTSDSKAAALATATK	293	

[illegible]

RESULT 4

hypothetical 76k protein - Chlamydochila pneumoniae (strain AR39)
C:Species: Chlamydochila pneumoniae, Chlamydia pneumoniae
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 31-Mar-2000
C:Accession: I40729
R:Perez-Melgosa, M.; Kuo, C.
Infect. Immun. 62, 880-886, 1994
A:Title: Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76-kDa
:Reference number:I40729; MUID:94156481; PMID:7509320

A:Accession: I40729
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-715 <RES>
A:Cross-references: GB:L23921; MID:g435961; PIDN:AAA23117.1; PID:g435962
A:Experimental source: Strain AR-39
A:Comment: This is the hypothetical translation of a sequence that was reported as two s

Query Match	14.18; Score 453; DB 2; Length 715;
-------------	-------------------------------------

Best Local Similarity 28.9%; Pred. No. 4, 2e-12;
Matches 136; Conservative 94; Mismatches 193; Indels 48; Gaps 14;

```

0y 11 GPESVSSNMOSSNPIIINOIASNSETKESTKA-----SEASB--SASSSVSSWSEFLSSAK 63
Db 262 GPGPIDETERPPADLSAOGLEASAAKSAEORJAGAEAKPKESKTSVERSWSLIRSAV 321
0y 64 NALISLRD -AIINKNSSPTDLSQLEASTSTSTVRAAKDYOBAKSNPTAKSGLEMA 121
Db 322 NALMSLADKLGIASSNSSSTSTRS-ADVDSTYATAPTPPPPEFDYKQAOATAYDTIFTS 380
0y 122 KTLAETRKMDLMAALODMERLANSDPSSNMHTEEVNNIKKALEAKOJTIID-----K 173
Db 381 TSLADIQALVSLQDAVINIKDTATD-----EET-----AIAAEWETKNAADVAKVGAQ 429
0y 174 LMKLVTLONOKSKLTEVUKTTDSADQIPAINSOLEIINKNSADQIITKDLERONI---SYEA 230
Db 430 ITELAKRYASDNOAIIIDSGKLTISFDLLQALLOSVAANNKKAELLEKEQODPVPVPGKTPA 489
0y 231 VITNAGEVIKASSEAGIRLGOALOSIVADGQSOAAVLOAOONNSPDNIATPFKELIDAAE 290
Db 490 IAGSLVDTDATATQIEKDGNAIRPAYFRAGONASSAVENTAKSNMSISINDSKAAIATAK 549
0y 291 TKVNEIKOEHTGLTDSPIVKRAEQISOAKNDIOEIKPS-GSDIPIVGP-SGSAASAGSA 348

```

Db 530 TQIAEAQKK --- FPDSPILQAEQWVIAQENDKLNKIPADGSDVPNPGTVGGSKQOQSS 6066

Qy 349 AGALKSSNNNGRISILLDDVDNEMAALALAGFSRMIQEPVNNVNPATAKELQAMEAQLTAM 4088

Db 607 IGSI - - - - - RVSLTLDDAENETASTILMSFSRMIHMFENFNENDSDSAAQOELAAQRAA 6599

Qy 409 SDQLVGADGELPAEIQAIKDALAQL - KQPSADSLATAMQDVAFPAAKVGG 458

Db 660 K - - - - - AAGDSSAAALADQKQKLEALGAGQOQOILNALGQIKSAVAVSAG 707

RESULT 5

F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cul, L.;
me, A.; Mizutani-Ol, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirumatsu, K.

Lancet 357, 1225-1240, 2001
A/Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*

A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F90073

A;Status: prelimina
A;Molecule type: DN

A;Residues: 1-2271 <KUR>
A;Cross-references: GB:BA000018; PID:g13702612; PIDN:BAB43752.1; GSPDB:GN00149

A;Experimental source: strain N315
C;Genetics:

A;Gene: SA2447

Query Match 8.4%; Score 268; DB 2; Length 2271;

Best Local Similarity 20.6%; Pred. No. 0.0013;
Matches 146; Conservative 132; Mismatches 271; Indels 160; Gaps 22

QY 10 SCPEVSNSQSMNPITINGQIASNEFTKESTKASEASPSASSVSSWFLSSAKNALISL 69

```

Db 1134 SCSYSTSTLSNERTSVSDSTLSLSESDSISTESTSTDSISAIASBESTSIS - 119

```

QY 70 RDAILINKSSPIDSLSQ-----LEASTSTSTVTRVAAKDYDEAKSNFDTAKSGLEN 120

```
Db 1193 -----LSESNSTSDSESQSAFLSESLSESTSESTSESVSSSTSESTSLDSETSSEG-ST 124
```

QY 121 AKTIAEYETKADLMALQ-----DMERLANDPSNNHTEEVYNNIKKALEAQKDTID 172

```
Db 1248 STSLNSTSCGASISTSTLSESTSTFKSESVSTLSMSTSLNSTSLSDSTSD 130
```

QY 173 KLNKLVTLQNGKSLTEVLKTTDS-----ADQIPAINSOLEINKNSADQIIKDL 221

```
Db 1308 -----SKSDSLSTMSKDSISTKSDSLSTLSGSGTSESESDSTSSSESKSDS 1335
```

QY 222 ERONISYE-----AVLTNAGEVIKASSEAGIKLGQAL-QSIVDAGDQSQAAVLQAGQNNNS 275

```

Db      1359  TSMISMSQSTSGTSTSTSTSLSDSTSTSLASASNQGVDSMASQASASNSTSTSTS 141
          :| : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 276 PDNIATKELIDAEETKVNELKQEHGTGLDSPLYKKAEEQI-----SQAQKDIQ 324

```
Db 1419 ESDQST-STVNSQSTSQSESTSTSTSLDSTSIKSTSQSGSTSTASLSGSESESDQ 147
```

QY 325 EIKPSGSDPIVGPSSGSAASAGALKSSNNSGRISLLDDVDNEMAAIALQGRSMI 384

```

Db 1478 SISTRASE-----STSESASTSLDSTSTNSGAS----- 150

```

QY 385 EQFNVNNPATAKELQAMEAQITAMSDQLVGADGELPAETQAIKDALAQALKQPSADGLAT 444

```

Db 1509 TTTSLNSASASASE--SDSSSTSLSDSTSTASMSQSSSEDSQSTASLASLSDLSTSTSNRMST 156
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

445 AMGQVAFAAAKVGGGAGTAGTVQMNVKQLYKTAFSTSSSSYAALSDGYSAVYKTLNSL 504

```

Db 1566 ----IASLSTSVSTSECS-----TSESTSESDNSTSLSDSQSTR----- 160

```

QY 505 YSESRSQVQSAISQTANPALSRVSRSQIESQGRSADASQ-----RAETIVRDSQTLGD 559

Db 1604 -STASGASASTSTSD---SRSTASASTSTMTSTSDOSMSLSTSTSTMSLSTSLD 1659
 QY 560 VVS-----RLOVLDs-----LMSTVSPANOELMOKLTLASI 593
 Db 1660 SVSDSTSDSTASSTSGMSVSIISLSDSTSTSTASSEVMSASISDSGMSSESVDSSEVSE 1719
 QY 594 SKA-----POGYPAVONS-----ADSLQKFAQLEREFGERSLAES 632
 Db 1720 SNESDSKMSGSGTSSVSDSGSLSVSTSLRKSESVSESSSLSGSGSMDS 1768

RESULT 6
 B89921
 hypothetical protein ebha [imported] - staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: B89921
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani, U.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsuku, K.
 A:Title: Whole genome sequencing of methicillin-resistant staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: B89921
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-6713 <KUR>
 A:Cross-references: GB:BA000018; PID:g13701232; PIDN:BA842527.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics: ebha

Query Match 8.1%; Score 261; DB 2; Length 6713;
 Best Local Similarity 20.5%; Pred. No. 0.012;
 Matches 150; Conservative 133; Mismatches 237; Indels 210; Gaps 31;

QY 10 GSPESVSSNOSSMNPITINGOIASNETKSTKASESPASSSVSSFLSSAKNALISL 69
 Db 5610 NGDQNLANAKDKANAFVNSLNGLMQOQO-----DLAKHKAINNADTVSDVTDIYVNGIDL 5663
 QY 70 RDAILKNSSPTDSLSL---EASTSTSTYTRAAKYDYDAKSNEFDPAKSGLEAKTILAE 126
 Db 5664 NDAM-----ETLKLHLYVNEIPNAQETVYVNOAD--DNAKTNFEDAK-----5702
 QY 127 YETRMADLALADMERLANSNNTTEENIKKALEOKDTIDKLN-----KLV 178
 Db 5703 -----RLATYTLNSDMT--NVNDINGALQAVNDAIHLNGDQRLQDADK 5745
 QY 179 TLONQNKSLTEVLKTTDSA-----DQIPAINSOLE-----INKNSADQIINKLERQ- 224
 Db 5746 AIGSINGALANKLKEIRASVATDQDKLIANKKAEELANSIININKATSNQAVSOVOTAG 5805
 QY 225 NISYEAVLTNAGEYTRKASSERGIKLGALOSIYDAGOSQAVYQAOQNNSPNIAATKE 284
 Db 5806 NHALEYQV--HANETPKAKIDANKVDQVQALIDEIDR-----NPNLTDK 5848
 QY 285 LIDAEKRVNE-LKOEHTGLTDS---PLYKKAERQISOAQKIDQIEKPSGSDIPYVPSG 340
 Db 5849 EKQALKRINQIIOGGHNDINNALTKEIEQAKQALQALQIDQIDKLVKAKED-----5900
 QY 341 SAASAGSAGALKSSNNGSRISLLDDVDN--EMAIALQGFMSIMEQF-----NVNPN 392
 Db 5901 -----AKQVDQKQVQALIDEIDQNPNLTKREKQALKDRINQIIOGGHNGINNA 5948
 QY 393 ATAKELQAMEQLT-----AMSQDLYVAGDELPAEIDAIK 427
 Db 5949 MTKKEIEQAKQALQALQALKEIKDYKAKENAKQVDQKQVQALIDEIDQNPNLTKREKQALK 6008
 QY 428 DALAQALQKPSADGLATAMQOVAFAAKVGGSAGTAGTYOMNKKLYKTAFSTSSSY 487
 Db 6009 DRINQIIOGGHND--INNAMTKKEIEQAK-----AQALQALQ--DIKDLVK-----6050

QY 488 AALSDGYSAKYTLNLSYSESGVQSAISQTPANPALSRVSVSGIESQGRASDAQORA 547
 Db 6051 --AKEDAKNAIK--ALANAKRDQINS-----NPDLPQAKAKAL-----KEIDEAEKRA 6095
 QY 548 ETIVRSQTLGDYYSRLQV--LDSLMSTIV-----SNQANQ-----EELM-----586
 Db 6096 LQVNEAQTIDQNLNGLNGLDRIHNVHWEVDQPAVNIIFPATPQIILVNGELLYHRD 6155
 QY 587 -----OKLTASISKAPQFG-----YPAVONSADSLQKFAQLEREFGERSLAESQENA 636
 Db 6156 DIITEQDILAHNLIDQLSFEVIDTPTATISLSL---TAKVEVTLIDDSKVYVNPVKV 6212
 QY 637 FRQPAFIQ 646
 Db 6213 VEKELSVKQ 6222

RESULT 7

T12796
 probable transglycosylase - Bacillus subtilis phage SPBC2
 C:Species: Bacillus subtilis phage SPBC2
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
 C:Accession: T12796; A69911
 R:Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.
 submitted to the EMBL Data Library, August 1997
 A:Description: The complete nucleotide sequence of the Bacillus subtilis spetac2 pro

A:Reference number: 217583
 A:Accession: T12796
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2285 <LA2>
 A:Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025510; PIDN:AAC13005.1
 R:Kunz, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allioni, G.; Azevedo, V.; Ber C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.D.; Emmerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal lech, J.; Harwood, C.R.; Henalt, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapius, A.; Lardino A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau Y, M.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl A:Authors: Schleich, S.; Schroeter, R.; Scrofano, F.; Sekiguchi, J.; Sekowska, A.; Se akeuchi, M.; Yamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: A69911
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2285 <KUN>
 A:Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14053.1; PID:e11835
 A:Experimental source: strain 168
 C:Genetics: yomi

Query Match 7.3%; Score 235; DB 2; Length 2285;
 Best Local Similarity 19.3%; Pred. No. 0.034;
 Matches 159; Conservative 141; Mismatches 312; Indels 210; Gaps 33;

QY 10 GSPESVSSNOSSMNPIT--NGQIASNSE-----TKSTASASPSASSSVSSWSFLSS 61
 Db 56 SAIDTYQKNLSKYNOTVETSTVKNADGSEVKLTQYKKNKGETLQRETKIINN-----109
 QY 62 AKNALISLRALINKNSPDSLSQLEASTSTSTVTRVAAKDKDEAKSNPD-----TAK 115
 Db 110 -RNTALEQTOEVNKLQATEKLGQYQKTYQNRNLQGPRTKYVQKNHNGDDIYVTTPDK 168
 QY 116 SGLENAKTLAEYETKMDLMAALQDMERL-----ANSPSNNHTEEVN 158
 Db 169 TNSTSKTTNTYNDQQRRAIEQLKQDEKLRQOGIYVDTTSSLGKRIKNTAQSAQOIEALQ 228

```

QY 159 NIKKALEAOKDTIDKLNKL-----VTLQONKSLFEVLKTTDS---ADQIPAI 203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 229 NRIIMLDDSSAAVAKNNEKKTIELYQROAVNQNILNTRYGSSMSSNQAVQDLYNAV 288
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 NSQLEINKNSADQIINKLERONISTEAVLTNAGEVITKASSEAGIKIGQALQSI---VDAG 260
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 289 NS-LNVSTGS-NNIRSOISLQNMQFRELASNAQFANQAQSSFGAELTQFFKSMSTYLLISG 346
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 261 D-----QSOAAVLQAQONNSPDNIATFKELIDAEKTVKVELQOEHLGLTSPDLP 308
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 SLEFGAISGLKEVNSQAIEIDTLMTN-----IRVNMPEPKYKVELLOQESIDLDGT-L 398
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 VKKA-----EQOISQAOKDIOEIKPSGSDIPVPGSGSAASGSAAGAL 352
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 399 SKRTIDILQMTGDFGRMGFDESELSLTKTAQVLA---NVSDLTPDDVNTLTAAMLNP 454
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 353 K--SSNNSGISLLDDVDNEMAIALQ-----GFRSKMT----- 384
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 NTAANDSISIADKLNEVDNVAVTTLDLANSIRKAGSTASTEGVELNDLIGYTTAIASTT 514
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 385 -EOFNV-----NNPATAKELOAMEAQLTAMSDOLVGADGELPAEIOAIKDAL 430
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 515 RESGNIVGSLKTIPTARIGNNOSSIKALEQIGISVKTAGEAKSA--SDLISEVAGKWDTL 573
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 AQALQKQPSADGLA-----TAM--GOVAFPAAKVGGSGAGT-----AGTVQMN 471
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 574 SQAQKQNTSIGVAGIYQLSRFNAMNMFSLQAQAAKTAANTSGSANSSEQOKVADSLQAV 633
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 472 KQLYK--TAFSTSSSSVYAAALSDGYSAY--KTLNLSYSESRGVSQ-----AISQTA 520
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 634 NKLQNNFEFAIQAASDAF---ISDGLIEFTQAAGSLNASTGVIKSVGLPLLAAVSTA 690
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 521 NPALSR-----SVSRGIESQGRSADASQRAETIVRD---SQTIG 558
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 691 TLLSKNTNTTASSLLILGTAMQOETLATAGLEAGMTRAAYASRYLKTALRGILVSTLVG 750
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 559 DVYSRL-OVLDSLMTSTIVSNPQANOEIMQKLTASISKAPQFCYPAVQNSADSLQFAAQ 617
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 751 GAFALGMALESILISFAAKKAKKD-----FEGSQQTNVEAITTNKSTDLIIQ 801
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 618 L-EREVDGERSLAESQENAFKQPAFIQOVLVNITASLFSGY 658
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 802 YKELQVKESRSLTSDDEQVLAQ---VTQOLAQTFFPALVKGY 840
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8

```

Db6731
Hypothetical protein y1hd [Imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: D66731
R:Boletijn, A.; Winker, P.; Mager, S.; Jallion, O.; Malarne, K.; Weissenbach, J.; Emrl,
  Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
  A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: D66731
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1063 <STO>
A:Cross-references: GB:AE005176; PID:g12723779; PIDN:AAK04950.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: y1hd

```

```

Query Match 7.3%; Score 234; DB 2; Length 1063;
Best Local Similarity 22.4%; Pred. No. 0.013;
Matches 160; Conservative 108; Mismatches 285; Indels 160; Gaps 29;

```

```

QY 13 EVSSNOSNMPI-----INGOIASNEFKESTKASEAS--PSASSVSSMSFL-- 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 DSSSSNDNSNLSLSSNNAADVSADVSGSSSTSSGVLSESSALDSGIAVSQSEEMNL 205
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 -----SSAKNALISLDAILNKNSSPTDSLQ-----LEASTSTSTVTRVAAKD-YDEAK 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 206 VGNSSASASSAAVASFPAITLATPMSVPMULTQALAAAPATISGSAILNTTGLDLYNQAI 265
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 109 SNEDFAKSGLENA-KTLAEYERK-MADLMALQDMERLANSDPNSNHTTEVNNIKKALEA 166
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 ST--VGISGLANIFSTLGFENIPGMTTAAALANGVQIYNI-----VGNIOEA--- 311
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 OKDTIDKLNKLVTLOQONKSLFEVLKTTDSADQIPAINSQLINKNSADQIINKLERONI 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 -----AANPGAFLENIKS-----AGLDVSOIPLVGOIAAFAFAT-----PSM 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 227 STEAVLT--NAGEVITKASSEAGIKIGQALQSIYVADQSOAAVLQAQONNSPDNI--AAT 282
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 SPAAMLTFLFENPTIPGLSSIPGASL--VLSPVLSAISTVTSGLVN-QLMTTSSNALGVN 407
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 KELIDAEKTVKVELQOEHLGLTDSPLVYKAERQISQAOKDIOEIKPSGSDIPVPGSG-- 340
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 FDLDTLVSLQGDNDLVNYLAGLVNNSAINRV-GQIAWSQ-----LSPPTISNIPLVGTTVNN 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 341 -----SAASAGSAGALKSSNNGRISILLDDVDNEMAIALQGRSMIEQFNV-- 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 VLSPTLNLITGASLGEVA-----NLTVSSILLDOVNNISLGNLISLGTALATTENTIQ 514
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 390 -----NNPATAKE-----LOAMEAQLTAMSDOLVGADGELPAEIOAIKDALQALQK 436
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 515 NSLNSFGNLPAGASDILNQLVNLQALNAINNIVESATGIYNNLPG-LGAIENGSLNTISQI 573
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 437 -----PSADGLATAMQOVA-----FAAKVGGSGAGTACTVOM 469
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 574 PNINNFVNNAALNGITTTIINSTLPSVGASTVPNNSANSOSSSSSAASSSTSSS 633
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 470 NYKQLYKT-----AFSTSSSSVYAAALSDGYSAYKTLNLSYSESRGVSQASISQTA 523
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 634 NWSNNTSSNSSEANNTSSSTSNMASSSSSEGSAA-----SSNSSESSVYASSSDVSQS 688
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 524 LSRVSRGIESQGRSADASQRAETIVRDQSLDGVYSRLQVLDSLMTSTIVSNPQANOE 583
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 689 SSAGVNSSSSSAGSGASSSSNSSESVASSS-----VDSQS--SSAGVNS 734
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 584 EIMQKLTASISKAPQFCYPAVQNSADSLQFAAQLEREVDGERSLAESQENA 636
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 735 SSSSESSASSSSNSSESVASSSDVSQSSAGVNSSSSSAGSGASSSSNS 787
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9

```

T30822
Imp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999
C:Accession: T30822
R:Jensen, L.T.; Ladefoged, S.; Birkeby, S.; Christiansen, G.
  Infect. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis Pg21 deletion mutants by cultivation in the
  A:Reference number: 218884; MUID:95369882; PMID:7543881
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: EMBL:U21962; NID:g790243; PID:g790244; PIDN:AAA81013.1
A:Genetics:
A:Gene: Imp1
A:Genetic code: SGC3

```

```

Query Match 7.2%; Score 231; DB 2; Length 1365;
Best Local Similarity 21.5%; Pred. No. 0.023;
Matches 149; Conservative 137; Mismatches 288; Indels 120; Gaps 28;

```

```

QY 14 SVSSNOSNMPIINGOIASNSFKESTKASEASPSASSVSSMSFLSSAKN-----ALIS 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 SMQSAKSSLDAAV-AEITKKELETFNKDEAKNELKOTRNQIQEFTINTKKNPNSELIS 304
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 LRDAILNKNSSPTDSLQLEASTSTSTVTRVAAK-DYDE-----AKSNFTAKSGLENA 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 305 QLTSKRDSKNSVTDSSNKSNDIESANTELKQALAKANADVQADNLAKSKEIQLNINSVNA 364
Qy 122 KTL-AEYETKMAIDLMAALODMER---LANSDPNNHTTEEVNNIKKALEAKODTIDKLNK 177
Db 365 NTLISALTLTKDNTIQAKTELEKEVOKAQQAIKSNNTASQSAKSSLDKAVAITR-KL 422
Qy 178 VTLQONKSLTEVLTQTTSDADQIPAINSOLEINKNSAD--QIKLERONISYEAVL-T 233
Db 423 ETEFNKKEKKEFNELKQTRN-----QIOEFINTNKNPNPNSSELLISQLSKRSKNSVTDSS 477
Qy 234 NGEVTKASSEAGIKLGOALQSTIVDAGSOAAVLAQONNSPDNI-AATKELIDAEFK 292
Db 478 NMSDISANTELKQALAKANADVQADNLAKS--IKELQNNSVSNMNTLSAKLTD-----K 531
Qy 293 VNELKOEHTGLTDSPLVYKRAEEOISOAKDIOEIKPSGSDPIVYSGSAAASAGSAGAL 352
Db 532 DNTIQAKTEL-----EKEVQKADQAIK-----SNTASQSAKSSL 568
Qy 353 KSSNNSGRISLLIDVDNEMAAIALOGFRSMTEQF---NVNPNPAKELQAMEAQLTAMS 409
Db 569 DAKVAEITKLETFNKKDEKKEFNELKQTRNOQIOEFINTNKNPNP-----NYSSELLISQLSKR 624
Qy 410 DQVLG-ADGELPAEIOAIDALQALQPSADGLATAMGVAFNAKVGGSAGTAGTYQ 468
Db 625 DSKNSVTDSSNKSNDIESANTELKQALAKANADVQA--DNLAKSIREQLNNSVSNMNTLS 662
Qy 469 MNV-----KQLYK-----TAFSTSSSYAALSDGYSAYKTLN-- 502
Db 683 AKLTQDNTIQAKTELEKEVOKAQQAIKSNNTASQSAKSSLDKAVAITRKLTFNPKD 742
Qy 503 -----SLYSESRGVSQSAISQTA-NPALSRVSRSQIESQGRSADASQRAAETIVDSQT 556
Db 743 KEAKFNEMLKQTRNOQIOEFINTNKNPNPNSSELLISQI-----TSKRDSKNSVTDSSN 792
Qy 557 LQGVYRLQVLDLSMTIYSNPQANOEIMQKLTASISKAPQPGYAVONASADSLQKFA 616
Db 793 KSDIESANTELKQALNTAKAKKSSINDLRLPKNDQSLKEERG--PIRNT--NFSWISS 848
Qy 617 QLEREFVDEGRSLAE--SOENAFRQOPAFIOQVYL 648
Db 849 KLETT-----KNKLAEEITKADAIKKNPSSSKQAL 878

```

```

RESULT 10
F84730
probable myosin heavy chain [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence-revision 02-Feb-2001 #text-change 02-Feb-2001
C:Accession: F84730
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: F84730
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1269 <SNO>
A:Cross-references: GB:AE002093; NID:g6598483; PIDN:AAC69932.2; GSPDB:GN00139
C:Genetics:
A:Gene: Atg932240
A:Map position: 2

```

```

Query Match 7.1%; Score 226.5; DB 2; Length 1269;
Best Local Similarity 20.8%; Pred. No. 0.035;
Matches 145; Conservative 142; Mismatches 286; Indels 125; Gaps 28;
Qy 8 MEGSPSVSSNOSMNP-----INGOIASNETKESKTRASPASPSSASSVSSWSPLSS 61
Db 242 LKTKSKAKEMEKMAASLQOEIKELNKKSENKVEALAKSSAGELAAVVE-----LAL 296
Qy 62 AKNAL-----ISLRDAIINKNSPPTDLSQLDEASTSTSTVTREAAKADYDEAKSNFDJAK 115

```

```

Db 297 SKSRLETEQKQVSTALIDE---LTQLEQKQKASERREKELSVLQDDAQTKGQAKL 353
Qy 116 SGLENAKTLAEYETKMAIDLMAAL-ODMERLANSDPNNHTTEEVNNIKKALEAKODTIDK 174
Db 354 SEDEGINSLAEELKELLESISKQDEKLR--ANEKLAELKKEKALEA--NVAEY 409
Qy 175 NKLVTLOONKSLTEVLTQTTD-SADQIPAINSOLEINKNSADQIINDLERONISYEAVLT 233
Db 410 SNVATVTEVCNNELEKLTSDENFSKTDALLSQALSNNSLEQKLSLEELH-----S 462
Qy 234 NGEVTKASSEAGIKLGOALQSTIVDAGSOAAVLO-----AQONNS-----PDNIAT 282
Db 463 EAGSAAAATQKNLELEDVVRSSQAEEKSOIKLEKTFNAEOKNMLEEQQLNLQI 522
Qy 283 KELIDAEFKVNELEKOEHTGLTDSPLVYKRAEEOISOAKDIOEIKPSGSDPIVYSGS 342
Db 523 KS--SDAEELKLSKSSSFLQTA--IEVAEEKKATITQMDIYKQKASEL----- 569
Qy 343 ASAGSAGALKSSNNSGRISLLIDVDNEMAAIALOGFRSMTEQFNVNPNPAKELQAME 402
Db 570 -----ELSLTQSSARNSELEEDL-----RIALQGAHEHEDRANTTHQ-----RSIE 610
Qy 403 AQITMSDQLVGADG-----LPAE---IQAIKDALQALK---QPSADGLATANG 447
Db 611 LKGLCOSQSKHEDAGGRUKDELTLQTEKYRIQLEDEQVSSLEKKGHGETADSKG-YLG 669
Qy 448 QVA-----FAAAVGGSGSAGTAGYOM-NVKKQYKAFSTSSSYAALSDGYSAYKTL 501
Db 670 QVAELQSTLEAFQVKKSSLEALNLTENKELTELMNVTSKKKLEATVDIYSV----- 725
Qy 502 NSLYSESRGVSQA-----ISQTPALSRVSRSQIESQGRSADASQRAAETIVDSQT 556
Db 726 --KISESENLEESIRNELNLTQKLESIENDLKAQIQ--ESEVEXKLKSAEESLQKGRE 782
Qy 557 LQGVYRLQVLDLSMTIYSNPQANOEIMQKLTASISKAPQPGYAVONASADSLQKFA 616
Db 783 IDEATTKRMELEALHQSLSIDSEHRLQKAMEEPTSRDSE-----ASSLTERKUR 830
Qy 617 QLEREFVDEGRSLAE--SOENAFRQOPAFIOQVYNVNAS 653
Db 831 DLEGIKITYEQLAEASGKSSSLKEK--LEQITGRALAA 866

```

```

RESULT 11
JG6009
surface-located membrane protein Imp3 precursor - Mycoplasma hominis
C:Species: Mycoplasma hominis
C>Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 21-Jul-2000
C:Accession: JG6009
R:Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
J. Bacteriol. 178, 2775-2784, 1996
A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis Imp gene sys
A:Reference number: JG6009; MUID:96213016; PMID:8631664
A:Accession: JG6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CAA64858.1; PID:g1197336
C:Genetics:
A:Gene: Imp3
A:Genetic code: SGC3
C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homolo
C:Keywords: duplication; membrane protein
F:1-24/Domain: signal sequence #status predicted <Sig>
F:25-1302/Product: surface-located membrane protein Imp3 #status predicted <Mat>
F:957-992/Domain: tetratricopeptide repeat homology <TT1>
F:993-1026/Domain: tetratricopeptide repeat homology <TT2>
F:1089-1120/Domain: tetratricopeptide repeat homology <TT3>
F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>
Query Match 7.0%; Score 225.5; DB 1; Length 1302;
Best Local Similarity 21.6%; Pred. No. 0.04;
Matches 150; Conservative 139; Mismatches 223; Indels 181; Gaps 38;

```


RESULT 13

T18296

myosin heavy chain - Entamoeba histolytica

C:Species: Entamoeba histolytica

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000

C:Accession: T18296

R:Guillen, N.

submitted to the EMBL Data Library, February 1997

A:Reference number: Z18865

A:Accession: T18296

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2139 <GUI>

A:Cross-references: EMBL:103534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1

C:Genetics:

A:Gene: mhca

C:Superfamily: myosin heavy chain; myosin motor domain homology

P:91-780/Domain: myosin motor domain homology <MMO>

Query Match 6.9%; Score 219.5; DB 2; Length 2139;

Best Local Similarity 20.9%; Pred. No. 0.14;

Matches 151; Conservative 120; Mismatches 272; Indels 179; Gaps 32;

QY 13 ESYSSNOSNMPIINOIASNSESTKSTKASEASPSASSVSWSF-----ISSAKNAL 66

Db 994 ESIDEKDEITK-LKGDIKLEEKKDLEODRADVSATKDDIAKKIKITTECEDADEI 1052

QY 67 ISLRDAI---LNKNSFTDLSQLE---ASTSTVTRVAAKYDEAKSNFTAKSGLEN 120

Db 1053 AKLEQLEDEENKNDLTJNLQOTQLKGETEKSLAQVAAT--KKASDRDRTLSQNLN 1110

QY 121 -----AKTLAEYETKMDLAALODMERLANSNHNTEVNNIKKALEAQKFTID 172

Db 1111 EKLTTRKLTAKADLEKIKIGLKODYEDLE-----DDKNIKEDGLRAQKIKELDEIT 1165

QY 173 K-----LNKLVTLONOKNSLTVLKTDSAD-----QIPAINSQL-----EI 209

Db 1166 KGADVSOYLQKQKEEYSQIAKMOEKEKALGNDYKNKEKTIKEKELEISQLEKIDETEV 1225

QY 210 NKSADQIIDL-----ERQNIYSEAVLNAGEVITASSAGIKKIQALQALSIYDAG 260

Db 1226 EKDEAKKKKEIKEMKALQEEKENESS--KNSTKDKKLEKDKQKLDMDMTAD 1282

QY 261 DQ---SQAAVLQAOQNSPDN---IAATKELID---AAETKYNELKOEHTGLT----- 304

Db 1283 NEKIKAKAKLEQALNENYONNHEKAVADALLNKKKQASOKELNSLKAELBALTKASVY 1342

QY 305 -----DSPLVKKA-EEQISOAQKDIOEIKPSGSDIPTVSGSAAASGASAGALKSSNNS 358

Db 1343 ESKKKDSENEKKAISEIIDQANEKLNIO---ADL-----RKATADLOEANEKKAIVEA 1393

QY 359 GRISLLID-----DYDN-----EMAAIALQGRSMIEQGRNVNMPATAKELQA 400

Db 1394 QORQLVADNKKMTTLEETIARDEENTYKVENYKVKRRKREADELEFANEMIDIEKKBRM 1453

QY 401 MEAQLTMSDOLVAGDELPAEIOAIKDALQALQPSADGLATAMQOVAFAAKVGGS 460

Db 1454 KEQOVKRLBELKETKDLKLAAT-AEKDSIFTAKKQSDAD----- 1492

QY 461 AGTAGVQNNVKKLYKTA-----FSSSTSSSYAAALSDGSAVKTLYNSLYSEKSGVO-- 513

Db 1493 -----LEELNKTVEEHDEVAKLNTQITKLRDQNSABEELNELSKADKKKKI 1542

QY 514 SAISOTANPLSLSSVSSTGIESQGRSADASQRAETIVRSQTLGDIYSLQVLDLSMT 573

Db 1543 SELEEQVNELESRPVGTG-----NADEN-----EIKIRDAQ-IADLNKALE-----MKG 1585

QY 574 IVSNP--QANOEIMOK---LTASISKAPQFGYPAVONSADSLQKFAQLEREFVGBRSL 629

Db 1586 VONNOLOATNKEKAKADNLTSKIE-----ITENEMKKLENAKRLLEDQKDEADKAV 1637

QY 630 AE 631
Db 1638 SE 1639

RESULT 14

T17451

fimbriae-associated protein FapI - Streptococcus parvaangulis

C:Species: Streptococcus parvaangulis

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T17451

R:Wu, H.; Fives-Taylor, P.

submitted to the EMBL Data Library, October 1998

A:Description: Nucleotide sequences of the fapI locus.

A:Reference number: Z18788

A:Accession: T17451

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2570 <WH>

A:Cross-references: EMBL:AF100426; NID:g3929311; PID:g3929312; PIDN:AAC9868.1

A:Experimental source: strain FW213

C:Genetics:

A:Gene: fapI

Query Match 6.8%; Score 218.5; DB 2; Length 2570;

Best Local Similarity 19.7%; Pred. No. 0.2;

Matches 140; Conservative 151; Mismatches 288; Indels 133; Gaps 28;

QY 13 ESYSSNOSNMPIINOIASNSESTKSTKASE-----ASPSASSVSWSFSSAKNALI 67

Db 70 ETVIAKETTLTTTDANVKKLSSENFSEKAEKITSLSQSESASESVS--ESISESVSY 127

QY 68 SLRDAIINK-----NSSPTDLSQ-LEASTSTVTRVAAKYDEAKSNFTAKSG 117

Db 128 STSESSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESV 187

QY 118 LENAKTLAEYETKMDLAALODM--ERLANSNHNTEVNNIKKALEAQKQITDKLN 175

Db 188 EKKODSVRENIDKRISEAEVNLNMAARKLITLD-----AEQLELMKSLVATQSOLEATK 242

QY 176 KLVTLQONOKNSLTVLKTPTSADQIPAINSQLIENKNSAQIILKDERQVISEAVITNA 235

Db 243 NLGDPNATVADIQIAYTTLGNNTQALGNELIKLNPNG--QITAVLNNTFASRAATL-- 297

QY 236 GEVYKASSEAGIKRQALQSIYDAGQSOAAVQAOQNSPDIATKELIDAEETKVN- 294

Db 298 -----RSTTGCTTTFTIFSPSGNGQYYWA--GNNANNKNKNISSAYYDATKISW 350

QY 295 ELKOEHTGLTDSPLVKKAEQIS--QAKDIOEIKPSGSDI-----PIVPGSNA-- 343

Db 351 TVEVDPTTLKSPALKTLKTYGTGYIDTSDSKLSTPTNVLIDGAATNPVTNFGNGSKG 410

QY 344 -----SAGSAGALKSSNN-----SGRISLLDDVDNEMMAL------QGRSMIEQF 387

Db 411 IEYVSKGTGKVTGKHTTTPDPAFSGRANDLADLEIKMLAATLSDPHFYEDGKNGRY 470

QY 388 N-----VN-----NPAKELQAMEALTA 407

Db 471 NGQTAPYVINDSGTALIGVQSVGNADSLPSDTTSGESASSESSTKIS--ESVIES 528

QY 408 MSDQVAGDELPAE--IQAIKDALQALQPSADGLATAMQOVAFAAKVGGSAGTAG 465

Db 529 ISSEVIGSVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESV 579

QY 466 TVQNNKQLYKTAFFSSSTSSSYAAALSDGSAKTKLYNSLYSEKRS--GYOSAIQOTANPAL 524

Db 580 SVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESV 637

QY 525 SRSVSRSGIE--SOGRSADSORAEIYVD--SOTGLADYVSRQVLDLSMTIV--GNPOA 580

Db 638 SESISESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESVSESV 696

QY 581 NOEIMOKLTASISKAPQFGYPAVONSADSLQKFAQLEREFVGBRSLAES 632

Db 697 VSESISESSESVSES-----ISESVSESSESISESSESV--SESISES 740

RESULT 15

E95206
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: E95206

R:Retellein, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
mon, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venier, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: E95206

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4776 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75846.1; PID:g14973269; GSPDB:GN00164; TIGR:SP4
C:Experimental source: strain TIGR4
C:Genetics:

A:Gene: SPI772

Query Match 6.8%: Score 218; DB 2; Length 4776;
Best Local Similarity 19.8%: Pred No. 0.48;
Matches 122; Conservative 109; Mismatches 363; Indels 22; Gaps 7;

QY 9 ESQPEVSSNQSMPDIINGQIANSSETEKASESPASSVSSWFLSAAKALIS 68

Db 2761 ESASTSASASTSASASTSASASTSASASTSASASTSASEASTSASASTSASASTS 2820

QY 69 LRDAILINKSSPDIQSLEASTSTVTRVAAKDYDEAKSNPDTKSGLENAKKTAEYE 128

Db 2821 ASASASTSASASTSASASTSASASTSASASTSASASTSASEASTSASEASTSASAS 2880

QY 129 TKMADLMALQDMERLANSPDNNHTEEVNINIKKALEAQDITDKLKVLTQNKKSIL 188

Db 2881 TSASEASTSASASTSASASTSASASTSASEASTSASEASTSASEASTSASEASTSAS 2940

QY 189 EVLKTDSADQIPAINSOLEINKNSADQIKLERONISTEAVLTNAGEVIKASSEAGIK 248

Db 2941 ESASTSASASTSASASTSASASTSASEASTSASEASTSASEASTSASEASTSASEASTSAS 3000

QY 249 L---GQALDSIYDAGQSOQAVALQAOONNSPDNIATKELIDAEFTKVMLEKQEHGLD 305

Db 3001 TSASASTSASASTSASASTSASTSASTSASTSASTSASTSASTSASTSASTSASTS 3060

QY 306 SPLVKAKEQISOAQKDIOETKPSGSDIPIVGPSSAASAGAAALKSSNNNGRISLL 365

Db 3061 ASASTSASEASTSASASTSASAS-----ASTSASASTSASTSASTSASEASTSASA 3115

QY 366 DDVDNEMAAIALQGRSMIEQFNVPNPATAKELQAMEAQLTAMSDOLVGADGELPAEIOA 425

Db 3116 SASTSASASTSASASA-----STSASASTSASASTSASTSASTSASTSASTSASA 3169

QY 426 IDDAALQALKOPADLAIAMGOVAFPAAKVGGSGAGT--AGTVQANVKOLYTAASSTS 483

Db 3170 SEASTSASASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASA 3229

QY 484 SSSYAAALDGYSAVYKTLNLSYSESSGVOSA---ISOTANPLLSVSSSGIEQGRSA 540

Db 3230 STSASASTSASASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSA 3289

QY 541 DASQRAETIVDSQTLGDIYSLQVLDLSIMSTIVSPQANOEIKQLTASISKAPQFG 600

Db 3290 SAS---ASTSASEASTSASASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 3346

QY 601 YPAVONSADSLQFAA 616

Db 3347 ASASTSASEASTSAS 3362

RESULT 16

T31110
extracellular matrix binding protein - Abiotrophia defectiva (fragment)
C:Species: Abiotrophia defectiva
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31110

R:Manganelli, R.; van de Rijn, I.

Infect. Immun. 67, 50-56, 1999
A:Title: Cloning and characterization of emb, a gene encoding the major adhesin of St
A:Reference number: 220988; MUID:99081722; PMID:9864195
A:Accession: T31110

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2055 <MAN>
A:Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AA03320.1
C:Genetics:

A:Gene: emb

Query Match 6.8%: Score 216.5; DB 2; Length 2055;
Best Local Similarity 21.3%: Pred. No. 0.18; Indels 263; Gaps 42;
Matches 178; Conservative 124; Mismatches 272;

QY 11 GPESVSSNQSMPDIINGQIANSSETEKASESPASSVSSWFLSAAKALIS 54

Db 501 GPETL-----PVAALRIYPRVNSGOVPEHRTSGNATNAFTIATVTPSTLND 552

QY 55 SMSFLSSAKN--ALISLRDAILINKSSPT-----DSLQLEASTSTSTVTRVAA 101

Db 553 QAISDATTIKSMVAVNDLEDARQTNCTPVQADIVYSKNGQVAAADIPVEGEYVTLRA 612

QY 102 KQVDEAKSNPDPAKSGLENAKKTAEYETKMADLMALQDMERLANSPDNNHTEEVNIN 159

Db 613 ROSQDKLSN-NTVQ--VKAFTLPREAK-----NAVNAAKAKNTAIDNNNLTAEEKA 664

QY 160 IKKA-LEAQDITDKLKVLTQNKKSIL-----ONKSLTEVLTDSADQIPAINSOLEINKN 212

Db 665 AEAKAYEAKKNATLADIDA-KTTAARNAQNKGTIDANVPVPAKRAANAAL--QA 721

QY 213 SADQIKLERONISTEAVLTNAGEVIKASSEAGIKLQAL--QSVYDAGQSOQAVALQOA 270

Db 722 AVNKRINEISORPDLTREERQAFMDQVRTARDAMAFAVSAANQAVTSARDQGLNVA 778

QY 271 QQNNSP-----DNIAATKE-----LIDAETKVN-ELK 297

Db 779 --NNLPTPAKPIPEALGHVROAADAKROAIRDANILTAEEQDALRQVPAAGTAAEAALN 836

QY 298 QERTGLTDSPLVKAKEQISOAQKDIOETKPSGSDIPIVGPSSAASAGSAGALKSSNN 357

Db 837 QNHTNAT---LAKAD---SDGKKAINDINPQ---PRSKPAANOALQEVYAAAKRQAALNN 885

QY 358 SGRISILLDDVDNEMA-AI-----ALQGRSMIEQFNVPNPAT-----394

Db 886 NNQLT-----DEEKQAQALQOVDQALANAKTOVQAANDNNNGVQAQTAQTAIINNINPOG 939

QY 395 ---AKELQAMEAQLTAMSDOLVG-----ADGELPAEIOAQLKQADLQKPSAD 440

Db 940 TOKAQAIITAEAEQAKRELQGRNDLTTEERNNALADLTAKAQAKKDVVQNRNTGVA 999

QY 441 GL-----ATANGVAFPA 453

Db 1000 GAKNDGVADQIOTINPAVYKTPDARNAIDQARKEAEFOANTKLTDEEKAAMAKKYQDAA 1059

QY 454 --AKVGGGAGTAGTVQANVKOLYKTAFSSTSSSTAAALSDGY-SAYTTLNLSYSESS 510

Db 1060 RDAKKAIDAGSGNDVNNVNNQ-GKAAIOAIK-----ALDSQPSAKTAKAALIONAAD 1112

QY 511 GVOASISOTANPALSHSVRSRGF---ESQGRSA---DASQRAET-IVRDS--OTLGDV 560

Db 1113 AKKAAI--TANNALTOEEKAAAIKQVDEDAKQAQAAVDSRSKADVDRAKDGQGLQISDV 1170

QY 561 -----YSRLQVLDLSMS---TIVSN---PDANOEIIMOKL-----589

Db 1947 DKTASLNLQTHIDLVHPHKKPKDAEKTINDDLARVATLQVNRKYSNNRKADALKAITATL 2006
Qy 466 TVQNMVVKLYTAFESTSSSS-----YAAALSDGYSAKYTLNLSYESSRGVSAISOAN 521
Db 2007 KLQMD--BELKAFARNADVDALKRFNVALSD-----IEAVITEKEN 2046
Qy 522 PALSRSVRSRGIESGRSADASORAETIVRDSOTLGDVYSRLQVLSLMTSTYSPNPA 581
Db 2047 SLKRDINAQOQTYAKFAIATPEQLAKYKVLIDQYVAD-GNNMIDEDATLNDIKQHTQFI 2105
Qy 562 QOEIMQ-KLTASISK-----APQFGYPAVQNSADSLQKFAQLEREFDVG 625
Db 2106 VDEILAIKLPATKATVPKPEIQAPKVCPTPIKKEETHESRKVEKELPMTGSSG 2158

RESULT 19
T13030
microtubule binding protein D-ClIP-190 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13030
R:Lantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A:Title: A class VI unconventional myosin is associated with a homologue of a microtubule
A:Reference number: 217588; MUID:98139549; PMID:9472041
A:Accession: T13030
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1690 <LAN>
A:Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AA896783.1
A:Experimental source: strain Oregon R
C:Genetics:
A:Cross-references: FlyBase:FBgn0020503
C:Keywords: cytoskeleton

Query Match 6.6%; Score 212; DB 2; Length 1690;
Best local similarity 19.3%; Pred. No. 0.21;
Matches 136; Conservative 133; Mismatches 300; Indels 134; Gaps 26;

Qy 4 HHHHESCPESVSSNOSSMNPITNGQIASNSE-----TKESTKASEASPSSSVSS- 55
Db 929 YHTDVESTKQLEAANMALEKYEASRAESDLODKVKEITPTLHAELQAEKSSSSA 988
Qy 56 -----WSFLSSAKNMLISLRDA-----ILNKNSPTDLSQLASHSTSTVTVA-- 101
Db 989 LHTKLSKSDSATGCHKELTSSKADWSQEMLOKELELOELRQLODSODSQKRLAEGER 1048
Qy 102 --KDYDEAKSNF--DTAKSGLENAKTLAEYETKMDLM-----AALQDMERLANSPS 150
Db 1049 KEKSEESIKNQEEVETAKTENLELSTGTOTTIKDLOERLEITMAELQHKEMASEDAQ 1108
Qy 151 -----NNHTEEVNNIKKALEAKQDITDKLNLVTLQONKNSLTEVL 191
Db 1109 KIADKLTVEAIOVANANISATNAELSTVLEVLQAEKSETNHIFELFEWADNMSERLIE 1168
Qy 192 KTTSDADQIPALNSQLELNKNSADQIILKDERONISYEAVLTVNAGEVIRKASEPAIKLQ 251
Db 1169 KYTGKEELKETHLQDRQKKFEELERLQAQOSEKLODES---QTSKEKLTETIQ 1224
Qy 252 ALOSIVDAGDSQAAVLQAOQNSPDNIATKELIDAETKKNELKQETHGTGLDPSLVK 311
Db 1225 SLQELQDSVKQKEELY-----QNLBEKYRESSITIEAONTKINESNOLEKNTS--CKE 1277
Qy 312 ABEQISQAKDIOETIKPSGSDPIPIVPSGSAASAGSAGALKSSNNSGRISLLDDVDNE 371
Db 1278 TODOLLESQKREKQLOEBAKL-----SGELOQVOEANGDIKDS--LVKVELVAVLEEK 1330
Qy 372 MAAL-----ALQFGSMIEQFVNNPATAKELQAMEAOLTAMSDQVAGADGLPAEIQAI 426
Db 1331 LQATTSQDQAQATNKELQELLVKSOENEGNIGESLAVTERKLOLEQANGEL----- 1383
Qy 427 KDALQALKOPSDADGLATAMGQVAFPAAKVGGSGAGTAGTVQNMVKQLYKTAFFSSTSSSS 486

Db 1384 KEALQ-----KENCKLELQGLK-----DESNTVLESQKK-----SHNE 1417
Qy 467 YAAALSDGYSAKYTLNLSYESSRGVSAISO--TANPALSRSVRSRG-IESQGRSADAS 543
Db 1418 IODKLEQAQOQKERTL-----QOETSKLAEQLSOLQKANEELKLSLOOKOLLEKGEFPTQ 1473
Qy 544 QRAAETIVRDSOTLGDVYSRLQVLSLMTSTI-----VSNPQANQ-----EETIMOKL- 589
Db 1474 LAEYQKVIDEMDDAASVSAL--LEQLONRYVAELETALRQANDQKTYLETKELRQLE 1531
Qy 590 TASISKAPQ-FCYPAVQNSADSLQKFAQLEREFDGERSLAE 631
Db 1532 SLELEKSRREVLSLKQNMGCASS--RSGKGDEVESLDIETSLAK 1572

RESULT 20
S33441
EF protein - Streptococcus suis
C:Species: Streptococcus suis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S33441
R:Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are abse
A:Reference number: S33441
A:Accession: S33441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1822 <SM1>
A:Cross-references: EMBL:X71880; NID:g298031; PID:CA50714.1; PID:g298032

Query Match 6.6%; Score 211.5; DB 2; Length 1822;
Best local similarity 19.6%; Pred. No. 0.24;
Matches 147; Conservative 131; Mismatches 300; Indels 171; Gaps 30;

Qy 15 VSSNOSMNPITNGQIASNSESTKAS-EASPSASSVSSWSFLSSAKNMLISLRDAI 73
Db 902 IAYNEDVINAQLODLNKLNDSEETKAAIDANPMLTPEEKA-----KAIAVEELY 953
Qy 74 LKNK---SSPTDLSQLEASTSTSTVTRV-----AAKDYDEAKSNF--DTAKSG 117
Db 954 NNAESDILSKPTPEVQAVEDKADKDLAKVELQAAADGAKGIEANPMLTPEEKDVAKKA 1013
Qy 118 LEN-----AKTLAEYETKMDLMAALQDM-EFLANSDPSNNHTEEVNNIKKL- 164
Db 1014 VEDAVKVAIDAIKASTPEVDFATSDGVKALDAEFKATQDAKKNKIAKEAESAKKAID 1073
Qy 165 -----EAQKDTIDKLNKLVTLQ-----NOKKSLTEVLKTTSDAQIPA 202
Db 1074 DNPMLTPEEKESAKNAVEBAKVATPAIDKASTPDAVOVEEDKGVAAINLITAKADAGV 1133
Qy 203 INSQLELNKNSADQIILKDERONISYEAV--LTNAGEVIRKASEAGIKGLQALQSIVD 258
Db 1134 IAAKL-----ADEIKLEDKQAEKAIIDASTMTNEEKATAK-----KALQDVVD 1178
Qy 259 AG-----DOSQAVALQAOQNSPDNIAT-----KELIDAE----- 290
Db 1179 KGKAELEDAARVATNEIEHATTEKAKAEELAGEKSLTDTGTEADAVELAKKELAKEA 1238
Qy 291 -----TKVNELKQENT--GLTDSP-----LVKKAEEQISO-----AQDI 323
Db 1239 IRTEEEETKIVKELAEPTKRAIENPNLSDDEKQAEIKKLLDVAVKLATIRNADAKRT 1298
Qy 324 QEIKPSGSDPIPIVPSGSAASAGSAA-GALKSSNNSGRISLLDDVDNEMAIALQGRS 382
Db 1299 QEAERKAQALADLEKAKETQIADKKAIDRLFTLVKGELEATKQDAKKIKAIADAARKEA 1358
Qy 383 MIEQFVNNPATAKELQAMEAOLTAMSDQVAGADGLPAEIQAITDALQALKOPSAGCL 442
Db 1359 IASNPNLTDAEKKTFTTDDVAVDEVARAKANDAISAATS--PADVQKEEDAGVAAIIEVDLDA 1416
Qy 443 -ATAMGQVA--FAAKVGGSGAGTAGTVQNMVKQLYKTAFFSSTSSSSYAAALSDGYSAVK 499

Db 1417 KODAKKRIKDAAAKAAIGS-----NPNLDAEKJFTTDAVDAEVAAK-NDATSAAT 1468
QY 500 TINSLSRSRGVS-----AISOTANPALSR--SVSRSGIESQGRSDASQRAETI 550
Db 1469 SPADVCKEEDAGYAAIAEDYDLAKODAKKIKKESDASAIIDANPNLTDARESKKA 1528
QY 551 VRDSOTLGDVYSRLQVLDLSMTSTIVSPQANDEIMOKLTASISKAPQEGYPAVONSADS 610
Db 1529 VD-----ADAKKATDAIDA--STSPVEAQSAEDKGYSIAQVDLDA-----AKQDAKK 1575
QY 611 LQKFAQLEREPYDGERSLAESQENAFK 639
Db 1576 IAKKVA-AAKEAIDANPNLSDAKKASKR 1603

RESULT 21

FmbB protein [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: D90011
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, T.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D90011
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2481 <KUR>
A:Cross-references: GB:BA000018; PID:g13701961; PIDN:BA943253.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: fmbB(mrp)

Query Match 6.6%; Score 210.5; DB 2; Length 2481;
Best Local Similarity 19.6%; Pred. No. 0.41;
Matches 153; Conservative 132; Mismatches 313; Indels 184; Gaps 32;

QY 19 QSSMNPILINQ-----IASNSEKESKASEASPSASSSSVSFLSSAKNALISLDAI 73
Db 659 QDVITPVPKQAKODITIAVTTTRKQOKKSMASLQDEKDVANDKIGKIEKAKIDDAAT 718
QY 74 LNK--NSSPDSLQLEASTSTSTVTRVAAKDYDE-----AKSNFDPAK-----SGLE 119
Db 719 TNQVRAIKTKAINDINOTTPATKAALAEEDVYQAIIDAPLNPDTTNEEVAATE 778
QY 120 NAKTLAEYETKMADLMAALQDMERLANSDBSNHTEEVNNIKKALEAQKDTIDKLKLVY 179
Db 779 RINAAKVSQKAIETATTAODLERKNEE-----IFKIENTDSTQCKMAYKEVRQAAT 833
QY 180 LQN-QKSLT-----EVLKT-----TDSAQIPAINSQLLEINKNSAQIITKDERON 225
Db 834 ARRAQNAVTVSNATDEEVAENAAVDAQAQTEGLDIOVYKQOEY---ADTKAKVLDKIN 889
QY 226 -ISYEAVLITNAGEV-----IKASSEAGIKGLQALOSIVDAGDSQAALQAOQ 272
Db 890 AITQQAQKVRKAADTEVENAYNTKQEIQNSNASTTEKEKAYITELDAKKQEAATNIDAN 949
QY 273 NNS-----PDNIAATKELIDAETKYNELKOEHT-----GLTDSPLVKAEEQT 316
Db 950 TNSDVTYAKNGIAAINOVQAATTKKSDAKAETIAQKASEKTAIEAMNDS-----TTEEQ 1005
QY 317 SQAKQIQEIKPESGDIPIYVPGSASASGASGALKSSNSGRISILLDDVDNEMAATA 376
Db 1006 AAKDKYQAAVVTANADI-----DNATANTDVNNAKTTEA--TIAITPDAVKKPA--A 1055
QY 377 LOGFRSMIE-----QFVNNPATAKELQAMEAOLTAUSDOLVGADGELPA-----ETQAIK 427
Db 1056 KQAIADKVAQOETAIDANNNGSTTEKEAAKQO---VQTEKTAADAALDAASHVNEVAEK 1112

QY 428 DALAQALK--QPS-----ADGLATAMQV-----AFPAKVGGSAGT-- 463
Db 1113 NAEIKTEAIOPATTTTKDAKQAIAATKANERKTAIAIOTODITAEETIAANADVADNVA 1172
QY 464 -----AGTYQMANKQLYKTAFFSTSSSSYAAALSDQYSAKYTINSLSYRSRQVQAIQ 518
Db 1173 NSNIEANSQNDVQAKTT--GETSIDQYTPYNNKATATKANETAILNKKLOEIQAATPA 1230
QY 519 T-----ANPALSRSVSRSGIESQGRSDASQRAAETIYRDSOTLGDVY 561
Db 1231 TDEKQAADAENATENGKANQAIATTAQVDEAKANNEAATINATPYVKKQAANKDEL 1290
QY 562 SRLQVLDLSMTSTIVSNPQ-----ANDEIMOKLTASISKAP----- 597
Db 1291 DQLQ-----ATQTNVINNDQATNNEKEBAIQIOLATAVTDANKNITTATDNGVDTAKDAG 1347
QY 598 ---QFGYF--AVQNSADSLQKFAQLEREVDCEBSLAESQENAFR-----KQAFIQ 646
Db 1348 NSIQSTQPAATVAVKSNKNEVDQAVTTQNOAIDMTTGATTEEKNAADLVAKAKAY-OD 1406
QY 647 VL 648
Db 1407 IL 1408

RESULT 22

T34107
hypothetical protein C18C4.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34107
R:Gatlung, S.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid C18C4.
A:Reference number: Z21478
A:Accession: T34107
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1091 <GAT>
A:Cross-references: EMBL:U55369; PIDN:AAC25825.1; GSPDB:GN00023; CESP:C18C4.5
A:Experimental source: strain Bristol N2; clone C18C4
C:Genetics:
A:Gene: CESP:C18C4.5
A:Map position: 5
A:Introns: 38/3; 95/3; 179/2; 259/3; 301/3; 419/2; 573/3; 613/1; 875/3; 920/2; 959/2;

Query Match 6.6%; Score 210; DB 2; Length 1091;
Best Local Similarity 19.0%; Pred. No. 0.14;
Matches 136; Conservative 141; Mismatches 289; Indels 150; Gaps 25;

QY 8 MESGPESVSSNOSMNPILINGQIASNSETEKESKASEASPS-----ASSVSSWSFLS 60
Db 108 MGEIENLQROLNITKTASIQSLMLAKSDSSKTKLSEBENTLTKLVEDIQKQVSPMSQ 167
QY 61 SAKNALI-SLRDAIILNKSSPTDSLQLEASTSTSTVTRVAAKDYDEAKSNFDPAKSGLE 119
Db 168 ODKNSEIQKKDAISYNDVSRQNMDSLSEKLSMDRTLREDOQKQSLQSQTETLKNAAS 227
QY 120 NAKT-----LAEYETKADL-----MAALQDMERLANSDBSN-- 152
Db 228 TSESTLSMLKDKLAQEQNALDLKNEAQAOKTSTRESILFESGRIRKLOALQALDSDENNA 287
QY 153 -----HTEEVNNIK-----KALEAQKDTIDKLKLVLYQ---NONKSLTEVLKT 193
Db 288 ILNVQLREKDGKIDRIQVLDLAESRAQAEEDVYRDMKEIITISKKDDSNLNLQDELRL 347
QY 194 TDSADQIPAINSQLLEINKNSAQIITKDERONTSYRAVLTNAGEVYKASSEAGIKGLQAL 253
Db 348 TEEKYQ-----QAKKIENLDETIKQOETQ-----IRDLGRSIDKAK 384
QY 254 QSIYVDAGDSQAANVLAQOQNSPDNI--AATKELIDAETKYNELKOEHTGLTDSPLVK 311
Db 385 ROLQKMSQORQNEVARQGEDSARSMEKATKEIKKLKSQV-QILOQOLE--QDELEQK 441

Db 630 IELSEVKIDELQSSIEVLSTKLEREYOSSNLQNEELKLSLRNFEELOADLMSKAKNEELE 689
Qy 123 TLAE-----YETKADLMAALQDMERLANSDEPNHTE--VN 158
Db 690 QOIESSREFSVITASKEMKLKMSSEKQSEMSTASLAFQEMOSTQADVASDVKY 749
Qy 159 NIKKALEAKQRTIDKLKLVTLQONKSLTEVLKTTDSADQIPAINSOLEIKNS---AD 215
Db 750 QVESLLENKEPELEENLURA-----NLKDSMDKTLIDQSOLELQOOSDLAD 797
Qy 216 QIKKLERONISYEAVLTNAGE-----VIKASEAGIKLQALQOSIYDAGD 261
Db 798 RLQEDLRISDARVQELINQVSELQIEVSSREFSVITANKEMQKIDSS--EAQISEMT 856
Qy 262 QSOAVLQAOQNSPDNIA--TKELIDAAET-----KYNELK- 297
Db 857 ASLTFQEMOSTRADAVASEDKVELESLLENKEPELEENLNLKNSNGKYLEQS 916
Qy 298 -----QEHRTGLTD-----SPLYKAEQISOAQKDQIKRPSGSDPIVPGSGSA 343
Db 917 QULDAQOQSDLTRQEDLRISDARVQELINQVSDQSELETARQDPTNAVQV--MEALKS 975
Qy 344 SAGSAGALKSS-----NNSGRISLLDDVDNEMAIALQGRSMIEQFNNNPATKEL 398
Db 976 EGGESEYBALRAELDAVQEKGRSSDLYTSLBCK-----IQELETAIESSTAENQKSKTI 1030
Qy 399 Q-----AMEA-----QLTAMSDQVLGADGELPAEIOAKDAL 430
Db 1031 QDFTKVSLESOICELKSQNEQMEITNLNMDQSEMSSQLESANAE----- 1079
Qy 431 AQALKOPASDGLATAMGOV-----AFAAKVGGSGAGTGTV-----QKN 470
Db 1080 -TELRTSAERTIDKLRGEYEKTKAMQOEHLAEIVAKIESRDVENADQAKHKEEDR 1138
Qy 471 VKQLYKAFSSSTSSSYAALSDGYSAYKTLNLSYSESRGY--QASISQTPALSR- 526
Db 1139 LQSVIDTLRTSQSTIEESQAKSE-----ELNSRIKELQASTIEFQKALADTENKQEKV 1192
Qy 527 -----SYSRGIESQGRSADASQRAAETIVRDSQTLGDIVSRLOVDS 569
Db 1193 ELEKVOBOMLNLVQAFEEKYSIRLEWNS--SLSNANKLELAEEALQSKENTIVTLES 1250
Qy 570 LMSITVSNQAOEIMQKLTASISKAPQFGYPAYVQNSADSLQKFAOLEREFVGERSL 629
Db 1251 RIETISQOPEARLEE-----ANWKSQAMQGLTQSLQSOROLGEMHEKMEASDRKV 1304
Qy 630 AESQENA 636
Db 1305 IEVEBOA 1311

RESULT 27

C89921
hypothetical protein ebh [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: C89921
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani, O.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: C89921
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-3890 <KUR>
A:Cross-references: GB:BA000018; PID:g13701233; PIDN:BA842528.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: ebh

Query Match 6.5%; Score 208; DB 2; Length 3890;
Best Local Similarity 20.0%; Pred. No. 0.96;
Matches 152; Conservative 123; Mismatches 286; Indels 200; Gaps 32;

Qy 6 HHMSEGPSSVSSNOSSMNPDIINGOIASNSEKREKASEASPSASSSVSSWFLSSAKNA 65
Db 2953 HSIDTQSTTGTQOSINAYNAKLTAAKRVQOINQVLAGSPYDQINTNTSAANOAKSD 3012
Qy 66 LISLDAILKNKSSPFDLSIQLEAS-----TSISTYTRAAKY----- 104
Db 3013 LDHARQALTPDKAPLQAKNQTOLEOSINOPTDTGTTASLNAVNOKLAARQKLTENOV 3072
Qy 105 -----DEAKSNPDTAKSGLENATLTAEYTKMADLMAALQDMERLANS 147
Db 3073 LINGPTVONDKVAENQAKQTLTARGL-----TL-----DROPLATTLTGASNL 3120
Qy 148 DPS--NNHTEEVNIRK--ALEAKQRTIDKLKLVTLQONKSLTEVLKTTDS--ADQIPA 202
Db 3121 NQAQONNFTQOINAAQNHAALETIKSNITALTMTKLDVADNNTIKSGQNTYDAPPA 3180
Qy 203 INSQLEINKSADQIITKLERONISYEAVLTNAGEYKASSEA-----GIKLG--QALQS 255
Db 3181 NKQAYDNANNAKGVIGETTFPTMDVNTVNOKAAVY--KSTKDALDGOONLQRAKTEATPA 3239
Qy 256 IVDAGDOQAAY--LQAOQNSPDNIAAT--KELIDAEYKYNELKO--EHTGLTDSPL 308
Db 3240 ITHASDLNQAKNALTOQVNSAQVAVNDIKOTTQSLTATATGKRGVANHNOVQSDN 3299
Qy 309 VKKAE--EOTSQAKDIOETKPSGSDPIVPGSGSAASAGSAGALKSSNNSGRISLL 364
Db 3300 YVNADTNKKNDVNNAYNANDIINGNAQHPVITPDDVNNALSNVTSKEHALNGEAKLNA 3359
Qy 365 LDDVNEMAAIALQGRFSMIEQFNNNPATKELQAMEQTLTAMSDQVLGAGELPAEQ 424
Db 3360 KOEANTFALGHNL-----NLNN-----VORONLOSOINGAR---QID 3392
Qy 425 AIKDALAQALKOPASDGLATAMGOVAFAPAAKVGSGSAGTAGVQNNVQOLYTAFSSSS 484
Db 3393 AV-----NTIKQ--NATNLNSAGNLROAVA-----DKQVKRT----- 3424
Qy 485 SSYAALSDGYSAVYTLNLSYSESRGYQASISQTPALSR-----RSVSR----- 531
Db 3425 EDYADADTAKQNAV--NSAVSSA-----ETIINGTANFTMSVDVNRATSAVTTKNALN 3477
Qy 532 GIESQGRSADASQRAAETIVR--DSQTLGDIVSRLOVDSL-----MSTIVSNP 578
Db 3478 GDEKLVQSTQDARAIDALPHLNNQAKADVSKINAASINAGVNTVKQOGTDLNTRMGNL 3537
Qy 579 QA--NOEIMQKLTASISKAPQFGYP-----AVQNSADSLQKFAOLE----- 619
Db 3538 QCAINDEQ-----TTLNQNQYODATPFSKRTATYNNAVQAKDLINKSNGQNKTKQVTEAM 3592
Qy 620 -----REFVDGERSLAESQENAFRKPAPFIOQVLVNIASL 654
Db 3593 NOVNSAKNNLDGTRILLDQAKQTA-----KQOLNNMTHL 3625

RESULT 28

C82206
methyl-accepting chemotaxis protein VC1403 [imported] - Vibrio cholerae (strain N1696)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82206
R:Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers, L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: C82206
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-641 <HEI>


```

Db 1690 RELEE---ARASRDEIFAQSEKSEKKLGLEAIILOJEPFAASEARHAEQERELAD 1746
Oy 542 ASQRAAEITVRDSQTLGVYRSLOVLDSLMSTIVSNPQANOE---EIMOKLTASISKAQ 598
Db 1747 EIANAS---GKSALLDERKRRLEARIQLEBELEEOSENMELNFRFKTLQV----- 1797
Oy 599 FGYPVONSADSLQFAAOLEREFFVNGERSLAESOENAFRKO-----PAFIOQVLVN 650
Db 1798 -----DTLNSELA-----GERSAAOKSENA-RQOLEKONKELAKKLOLEGS 1838
Oy 651 IASLEFGYLS 660
Db 1839 VKSKFKATIS 1848

RESULT 30
T18351
Imp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C:Accession: T18351
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christensen, G.
Infect. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis Pg21 deletion mutants by cultivation in the pre
A:Reference number: 218884; MUID:95369882; PMID:7543881
A:Accession: T18351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1051 <JEN>
A:Cross-references: EMBL:U21961; NID:g790241; PID:g790242; PIDN:AAA81012.1
C:Genetics:
A:Gene: Imp1
A:Genetic code: SGC3

Query Match 6.4%; Score 204; DB 2; Length 1051;
Best Local Similarity 19.4%; Pred. No. 0.24;
Matches 139; Conservative 142; Mismatches 315; Indels 122; Gaps 30;

```

```

Oy 5 HHMHSGPSSVSSNOSSNMPI-----INGQIASNSETR---ESTKASASPSASSVSS 55
Db 144 YEHK---QNIENELNKYTPISLESTLIEIQATNNLILKLNSETEKNDINLNKKEQL 200
Oy 56 WSFLSSAKNAL--ISLRDAILNKNSSPTD-----SLSQLEASTSTSTVTRVAAKDYDEAKS 109
Db 201 KASISQANLLPOLSDNDSEIARAKKSLDAETKMANQAVASNTTASM-----QSAS 252
Oy 110 NEDTAKSGLENAKTLAEY---ETKMDLMAALQDMERLANSDDPSN-NHTEEVNNIKKAL 164
Db 253 SLDAKYA--ETTKLETFFNKDKAEKFNELKQTRNOIOEFINTNNKNPNYSELISQLSK 310
Oy 165 ENOKRTIOKLVTLQONKSLTEVLKTTSDADQIPAINSOLEINKNSADQIKDLEQ 224
Db 311 DSKNSVTSSNN--SDIESNTLQALAKA--NADQVQAN---LAKSIRKQLNNSVSA 364
Oy 225 NISYEAVLITNAGEVIRKASSEAGIKGLQALQSIYDADGQSOAVLQAOQNSPDNIATKE 284
Db 365 N-TLSAKLTDRKNTTO---QAkteLEKEIQKANQAIKSNMTSMOSAKSLAKVAELTK 420
Oy 285 LIDA---AETRVNELKOEHTGLTDSPLVKAEEQISQAQKDIQETKPSGSDIPVPSG 340
Db 421 KLETENKDKAEKFNELKQTRNOIOEFINTNNKNPNYSELISQLSKRDKNK---VTDS 477
Oy 341 SNAASGSAAGALKSSNNSRISLLDDVDNEMAAIALQCFRSMIOF---NVNPNATK 396
Db 478 NNSDIESATTELQALNTAKAK--KSSIDNELRPLK--NDLQSKIEFFGIRATNTFSWISS 534
Oy 397 ELQAMEAQLTAMSDQVGD---GELPAEIOAIKDALQALK--QPSADGLATAMGOVAF 451
Db 535 KLETTKNTL---AELTKDAIKNPNSSKQALKDSSQOVQVLGNELLTTIEEFGKVT 591
Oy 452 AAAYGGSGAGTAGTVQNM-----VKOLY-----KTAFFSTSSSSVAALSDGYS 496

```

```

Db 592 KNSNIGYRLFLKLAQEQFNNSDVDKLKNAMEEQKTLSSKKQKLGNOSTKDYLTQLSTEMS 651
Oy 497 AY-----KTLNLSYSESRGCVOSAIQOTANPLNSV-----SRSGIESGGRADASQRA 547
Db 652 TQESTIKRYIVNIQAIHRNNLNSQYRLQLEADKLIANNKRGYGDYGLIESLQKQDMLD--- 708
Oy 548 ETVVRDSQTLGVYRSLOVLDSLMSTIVSNPQAN-----OE 583
Db 709 DSVLSYDSDLSKDEPNK--ALRVLVGDYTKNPNVSSWFIKRRKRSIENTONLNLILYRPN 766
Oy 584 EIMOKLTASISKAPOFGYPVONSADSLQFAAOLEREFFVNGERSLA---ESQENAF 637
Db 767 EILDKRAKDLRAEKTIKFVDENINSIDQRAKRLKQETLANKNLSNFTLNHOKNOF 824

RESULT 31
MKW
myosin heavy chain B [similarity] - Caenorhabditis elegans
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Caenorhabditis elegans
C:Date: 13-Jun-1983 #sequence_revision 19-May-2000 #text_change 19-Apr-2002
C:Accession: T20770; T21629; A93958; A93287; A21074; A02992
R:Kershaw, J.
Submitted to the EMBL Data Library, November 1996
A:Reference number: 219322
A:Accession: T20770
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1963 <WIL>
A:Cross-references: EMBL:281499; PIDN:CAB04089.1; GSPDB:GNO0019; CESP:F11C3.3
A:Experimental source: clone F11C3
A:Accession: T21629
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1963 <W12>
A:Cross-references: EMBL:283107; PIDN:CAB0505.1; GSPDB:GNO0019; CESP:F11C3.3
A:Experimental source: clone F13A7
R:Karn, J.; Brenner, S.; Barnett, L.
Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983
A:Title: Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy
A:Reference number: A93958; MUID:83273600; PMID:6576334
A:Accession: A93958
A:Molecule type: DNA
A:Residues: 1-61, 'EMSVIQ', 65-376, 'V', 378-1963 <KAR>
A:Cross-references: GB:J01050; NID:g156399; PIDN:AAA28124.1; PID:g156400
R:Macchiann, A.D.; Karn, J.
Nature 299, 226-231, 1982
A:Title: Periodic charge distributions in the myosin rod amino acid sequence match cr
A:Reference number: A93287; MUID:82272395; PMID:7202124
A:Accession: A93287
A:Molecule type: DNA
A:Residues: 847-1333, 'R', 1335-1876, 'L', 1878-1963 <KCL>
R:Wills, N.; Gesteland, R.F.; Karn, J.; Barnett, L.; Bolten, S.; Waterston, R.H.
Cell 33, 575-583, 1983
A:Title: The genes sup-7 X and sup-5 III of Caenorhabditis elegans suppress amber non
A:Reference number: A21074; MUID:83232892; PMID:6571695
A:Accession: A21074
A:Molecule type: DNA
A:Residues: 1873-1963 <W13>
A:Cross-references: GB:V01494; GB:J01049; NID:g6783; PIDN:CMA24738.1; PID:g6784
C:Genetics:
A:Gene: unc-54; CESP:F11C3.3
A:Map position: 1
A:Intons: 21/3; 64/3; 111/3; 264/1; 525/3; 951/2; 1747/3; 1819/3; 1894/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle
F:174-175/Domain: myosin motor domain homology <MKOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:662-684/Region: actin binding #status predicted
F:766-780/Region: actin binding #status predicted
F:848-1963/Domain: coiled coil #status predicted
F:848-1162/Region: S2
F:1163-1963/Region: light meromyosin

```

F:125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: Arg (Lys) #status predicted
F:702,712/Active site: Cys #status predicted

Query Match 6.3%; Score 202.5; DB 1; Length 1963;
Best Local Similarity 21.9%; Pred. No. 0.65;
Matches 136; Conservative 112; Mismatches 244; Indels 129; Gaps 27;

```

QY 82 DSLQLASTSTSTVTVRVAKDIYDEAKSNFDTAK-----SGLENAKT-LAEYETKMDLMA 136
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 863 DKYKALEDSLAKKEKLR-----KELEESSAKLYEKTSLFTNLESTKQLDSAEERLAKLEA 919
QY 137 ALQDM-----ERLANDSPSNHTPEVNNIKKALEAQKDTIDKLNKLVTLQNMKSLT 188
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 920 QOKDAKQSLQELNDQADNE---DRTADVQRAKKIEAEVEALK-----QIODELMSELT 971
QY 189 EYLKTTDSAD-OIPALNSOLEIKNSADQIKDLERONISYEAVLTMAGEVIKASSEAG- 246
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 972 KAEEKQSDHQRSLQDEHQ-----QODEKIAKLKKEKKHQBETINRKLMEQLQSEEDKGN 1027
QY 247 -----IKLQALQSLIVDAGDSQ--AAVLQAOQNNSPDNIAATKELIDAAETK-----VN 294
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1028 HQNKVAKALEQTLDDLEDSLEREKRRARADLDKQKRVESGLKTAQENIDESGRQHDLEN 1087
QY 295 ELKQEHITGL-----TPSPLYKKAEBQISQAQKDIQETKPSGDIPIVPGSGSAASAG 346
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1088 NKKKKESELHSVSRLEDEQALYSKLQRIKDGOSRISELEE-----LENERQSRSKAD 1142
QY 347 SAAGALKSSNNSGRISLLDDVDNEMAAITAQ-----GFRSMIEQNVNPP- 392
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1143 RAKSDLQR-----ELEELQEKLDQGGATRAQVEYNNKREAEIATKLRDLEENMMHENO 1197
QY 393 --ATAKELQAMEAQLTAMSQDLVAGDELPAE-IOAIKDA--LAQALKQPSADLITAMG 447
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1198 LGLRKKHHDVAVELTDQDLNKKAKAYEKKAQAVRQAEMLAQDLDETSEKL----- 1252
QY 448 QVAFAAKKVGGSAGTAGYQVMNVKQLYTAPSTSSSTSSAAALSGTSAKYLNLNLYSE 507
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1253 -----NNEKLAQFELDTLTEL-----QSKADEQSRLQDFTSLKGRHLEENG 1295
QY 508 SRSQVQASISQRTANPALSRVSRSRGIESQGRSADASQRAETIVRSQTLGDVYSRLQYL 567
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1296 LVRQLEDAESQVNLTRLSQLTSLQLEARTRADEARERQYVAAQAKNYQHEAEQDQ-- 1353
QY 568 DSLMTSTVSNPQANQOEIMQKLTASISKAPQFGYPAYONSADSLQRFPAQLEREFPYDGR 627
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1354 ESLEEEI-----EGKNEILLROL-----SKA-----NAD-IQQMKARFE-----GEG 1388
QY 628 SL-AESQENAFKQPAFIQOV 647
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1389 LKKADELQDAKRROAKINEL 1409

```

RESULT 32

A59287
myosin heavy chain - fluke (Schistosoma mansoni) (strain Brazilian LE)

C:Species: Schistosoma mansoni

C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000

C:Accession: A59287

R:Weston, D.S.; Schmitz, J.; Kemp, M.; Kunz, W.

Mol. Biochem. Parasitol. 58, 161-164, 1993

A:Title: Cloning and sequence characterization of a complete myosin heavy chain cDNA from

A:Reference number: A59287; MUID:93211444; PMID:8459827

A:Accession: A59287

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1940 <WES>

A:Cross-references: GB:L01634; PIDN:AAA29905.1

A:Experimental source: strain Brazilian LE

C:Genetics:

A:Gene: MYH

C:Superfamily: myosin heavy chain; myosin motor domain homology

F:82-752/Domain: myosin motor domain homology <MMO>

Query Match 6.3%; Score 202; DB 2; Length 1940;
Best Local Similarity 20.4%; Pred. No. 0.67;
Matches 144; Conservative 116; Mismatches 271; Indels 176; Gaps 30;

```

QY 20 SSANPIINQOIASNSTEKSTKASESPSASSSVSWFLSAAKNALISIRDAIINKNS- 78
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 820 TVKQPMLN--IAQOE--EMKKAEE--ELAKLEEEYKLEKLKLEESQNVVLAQKKND 872
QY 79 -----SPDLSLQLEASTS-----TSVTYVRAAK----D 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 873 LPIQLQTEQDSLADAEKISKVLVLRGDMQRIKLEPERLADEQDAAMLTVEYKKMSAE 932
QY 104 YDEAKSNFPTASGLENAATLAETK-----MADLMAALQDMERLANSNHTPEV 157
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 933 IEELKQDVDELSLQKAE--QEKQTKDQNIIRLQSEMAQOQEMIGKLNKDKKN--LEEQ 988
QY 158 N-NIKKALQAKDTIDKLNKL-----VTIQNMKSLTEVLKTTDSADQIIPALNSOLEINK 211
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 989 NKRTQALQAEEDKVNHLKRLAKLESTIDEMEENIARQKIRGDYK--SKRKLEGLD 1045
QY 212 NSADQITKDLERONISYEAVLTMAGEVI-----KASSEAGI-----KLQALQSLIV 257
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1046 KATQETVDDLERYKRLPEQLRKKEAEIGLSKFEDEGLVAQLQRIKELQTRIQELE 1105
QY 258 DAGDSQAQAVLQAOQNNSPDNIAATKELIDAAETKYNELKQEHITGL--DSPLYKKAEBQ 315
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1106 EDLEAEARAAKSKAEKR-----QQLSELEEYVDRL-EPQDQATAQSDLTTRKKEAE 1156
QY 316 ISOAQKDQIEIKPSGSDIPIVPGSGSAASAGSAAKLSKSSNNSGRISLLDDVDNEMAAI 375
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1157 LMKLRDLQEDTR-----LQNEQALVATMRKKQSPAINELA-- 1190
QY 376 ALQGFMSIEQFVNNPATAKELQAMEAQLTAMSQDLVAGDELPAEIOAIRDALQALK 435
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1191 -----DQLDQANKAKAKAEKERSQPKALDQAHNV--DSLIMKAKLSKRTYVA- 1237
QY 436 QPSADGLATAMQVAPAAKVGGSAGTAGYQVMNVKQLYTAPSTSSSTSSAAALSDGY 495
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1238 -----LESQLEQSVYKLE-----ATRLNLN-EQASTKARSQSEVELQRLQEEAE 1281
QY 496 SAKTINSLXSESRVQSAISOTANPALSRVSRSRGISQGRSADASQRAETIVRSQ 555
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1282 SOLQNLNKTIKQQLSAQLEARHSLEDE--SRKAKANGVRRINTSLDLS-LRETTLEEQS 1338
QY 556 TLGDVYSRLQVLDLSMTSTVSN-----PQANQOEIMQKLTASISKAPQFGYPAYONSA 608
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1339 AKGDLDQRLQKLOGELQQLRSRGGGGVRSVEVEELKKNMAKI-----PALESFA 1390
QY 609 DS-----LQKFAQLERE---FYDGER-----SLAESQENAFK 639
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1391 ESAKSKGQLEKTKARLQGELEDLMDVVERANGLAQLERKQNNFNR 1437

```

RESULT 33

A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)

N:Contains: myosin ATPase (EC 3.6.4.1)

C:Species: Dictyostelium discoideum

C>Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 19-Apr-2002

C:Accession: A26655; A24728; S00250

R:Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986

A:Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium di

A:Reference number: A26655; MUID:87092266; PMID:3540939

A:Accession: A26655

A:Molecule type: DNA

A:Residues: 1-2116 <NAR>

A:Cross-references: GB:M14628; GB:M11938; NID:G167834; PIDN:AAA33227.1; PID:G167835

R:DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.

Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985

A:Reference number: A24728; MUID:86016788; PMID:3901008

A:Accession: A24728

A:Molecule type: mRNA
A:Residues: 2035-2116
R:Magle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostellium
A:Reference number: S00250; MUID:88112226; PMID:2828113
A:Accession: S00250
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1734-1893 <NAG>
C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolyase; nucleotide binding; P-loop; phos
F:1-818/Domain: globular head <HED>
F:89-747/Domain: myosin motor domain homology <MMOT>
F:179-186/Region: nucleotide-binding motif A (P-loop)
F:819-2116/Domain: alpha-helical rod <ROD>

Query Match 6.3%; Score 201; DB 2; Length 2116;
Best Local Similarity 20.7%; Pred. No. 0.83;
Matches 163; Conservative 130; Mismatches 313; Indels 182; Gaps 33;

Oy 9 ESGPSSVSSNOGSMNPDIINGOIASN--SETEKSTKAS--EASPSASSSVSSMSF--LSS 61
Db 756 EAREORISITIAIOATRGWIAKRYVYKQAREHTVAARITIOQLRAYIDFKSWPWKLFSS 815

Oy 62 AKNALISLND-----AIIKNSSPTDLSQ--LEASTSTVTRV-----A 100
Db 816 KARPILKRNFEKEIKEREILELKSNTDSTTQDKLEKSLK-DTESNVLDLQRLKA 874

Oy 101 AKD-----YD-----EAKSNFDTAKSGLENAKT-LAEYETKMDLMAA 137
Db 875 EKETLKAMWSDKALEAQKRELEIRVEDMESELEKILALENIQOKRSEVEKVRDLREE 934

Oy 138 LODMERLANS--DPSNNHTEEVNNTKKALEAKOTIDKLNKLV-ILQNNKSLTE----- 189
Db 935 LOEBOKLRATLEKLLKKKEELEEMKRVNDGSDTISRLEKIKDELQCKVEELTESFSEE 994

Oy 190 -----VLKTTDSADQIPAINSOLEINKNSAQDIIDLEQNISYEAVLTNAGEVIKASSE 244
Db 995 SKDKQVLETRVRLQSELDLTVRLDSETKDK--SELLRQKKKLEBELKQVDEALAEFLA 1052

Oy 245 AGI-----KLGQALQSIIVDAGD-----QSOAAVLAQOONNSPDNIAA 281
Db 1053 AKLAQEAANKKRLQGEYTELENEKFNESEVTARSNVEKSKTLESOVAVNNELDEKKNRDA 1112

Oy 282 TKELIDAAETKYNELKQ--EHTG-----LTDSPLYKKAE-----DQISAQGDIOEIKFS 329
Db 1113 LEKKKALDAMLEEMKDDOLESTGGERKSLYDLKVKQESDMALRNQDISLOSTIAKLEKI 1172

Oy 330 GSDIPYVPGSGSAA--SAGSAGALKSSN--NSGRISLILLDD-----VDNMAAIALOG 379
Db 1173 KSTL-----EGEVARLQGLELEBOLAKSNVEKQKKVELDLEKSAQLAEFLAAQALDK 1227

Oy 380 FRSMTIQ-----FNVNNPATAK-----ELOAMEAQLTAMSDQLVG 414
Db 1228 LKKKLEQELSEVQTOLESPANNNKVNNSDTNKHLETSFNNLKLELEAQKAXALEKKRLG 1287

Oy 415 ADGELPAEIQAIKDALQALQKPSAD-----GLATAMQV--AFPAAVVGGGSAGT 463
Db 1288 ----LESELKHYNEOLEEKKQKQESNEKKRVLEKEVSELKQDIEEVASKAATVEAKNK 1343

Oy 464 AGTVQNNVKOLKTPAFSSSTSSSYAALSDGYSAVKTTLNSTLSESGVQSAISOTAPNA 523
Db 1344 KESELDKIKROYADVYSSHDKS-----VEQLKTLQAKKEBELRNTAEELBQDLRAE 1394

Oy 524 LRSVSRSGIESQGRS-----ADASQRAAETIVRDSQT-LGD--VYSRLQVLDLS 569
Db 1395 RSKKRAEFLDEAVKNLEETAKKVKAKEMKAETDYSTKSELDDAKNVSEQVQLK 1454

Oy 570 LMSTIVSNQANQDELMOKLTASI--SKAPQGYPAVQNSASLSQKFAQLQEREFYDGER 627
Db 1455 RLNEBELSELRLVLEADERCNSAIIKAKTAESALSLAKDEIDAANNKAKAKERKSELEF 1514

Oy 628 SLAESQEN 635
Db 1515 RVAELEES 1522

RESULT 34
B90835
probable tail fiber protein [imported] - Escherichia coli (strain 0157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B90835
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gatawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90835
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA835073.1; PID:913361114; GSPDB:GN00154
A:Experimental source: strain 0157:H7, substrain RIMD 0509952
A:Gene: Ecs1650

Query Match 6.2%; Score 200; DB 2; Length 971;
Best Local Similarity 20.7%; Pred. No. 0.32;
Matches 121; Conservative 84; Mismatches 219; Indels 160; Gaps 17;

Oy 6 HIMESGPESVSSNOGSMNPDIINGOIASNSETEKSTKASEPASPASSSVSSWFLSSANNA 65
Db 144 HATDAADSARAASSTAGQAASSAQSSAGTATKATPASKASAAAESSKSAATSAAGA 203

Oy 66 LISLDAILNKSSPTDLSIQLEASTSTSTVTRVAAKDYDEKSNFDTAKSGLENAKTLA 125
Db 204 A-----KTSETNAVVSQOASATSTATTTKAS--EAASSARDAASAKAEAKS-- 248

Oy 126 EYETKMDLMAALQOMERLANSDPNNHTEEVNNTKKALEAKOTIDKLNKLVTLQNNK 185
Db 249 -SETSAAS-----SASSAASSATAGNSAKAA----- 274

Oy 186 SLTEVLKTTDSADQIPAINSOLEINKNSAQDIIDLEQNISYEAVLTNAGEVIKASSEA 245
Db 275 -----KTSE-----TNKSSSETAAEQ-----SASAAA 296

Oy 246 GIKLQALQSIYDADGSOAAVLAQOONNSPDNINATKELIDAAETKYNELKQERTGLTD 305
Db 297 GSKTAAALSASAASTASAGQASASATAGKSASBSAAS--ASTATTGAGEATEQASAAAS 353

Oy 306 SPLVKRAEQIOAQKDIOEIKPESGSDIPYVPGSGSAGSAGSAGALKSSNNSGRISILL 365
Db 354 SASAAKTSETNAKA-----SETSABESSKTAASASASASASASASASAK----- 397

Oy 366 DDVDNEMAIALQGRSMIEDPFNVPNNPATAKELQAMEAQLTAMSDQLVGADGELPAEIQ 425
Db 398 DEATRQASA-----AKSSAVTTASTKATB-----A 421

Oy 426 IKDALQALQKPSADGLATAMQVAFPAAKVGGGSA-GTAGTVQNNVKOLYTPAFSSSTS 484
Db 422 AGSATPAALQOSKSTASATRAETAKKRAEDIASAVALDEASTTKGIVQL--SSATNSTS 479

Oy 485 SSTAAALSDGYSAVYTTUNSLY--SESRSGVQSAISOTAPNALRSVSRSGIESQGRSADS 543
Db 480 ESLAATTPKAVKAYELANGKTYTAQDATTAAQKIYVLSNATNSTSEMLA-----ATRS 532

Oy 544 ORAAETIVRDSQTLGD-----VYSRLQVLDLSLMSTIVSNQ 580
Db 533 VAAAVDLANGKTYTAQDATTAAQGIYVLSATFNSASETTLAATPKA 576

RESULT 35
C85693

probable membrane protein of prophage CP-933X 2198 [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: C85693
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dialianta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85693
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-973 <STO>
A:Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UWGP:219
A:Experimental source: strain O157:H7, substrain ED1933
C:Genetics:
A:Gene: Z1918

Query Match 6.2%; Score 200; DB 2; Length 973;
Best Local Similarity 20.7%; Pred. No. 0.32;
Matches 121; Conservative 84; Mismatches 219; Indels 160; Gaps 17;

QY 6 HHMSGESEVSSNOSMNPIINGQAINSETKSTKASESPASSSSVSMFLSSAKNA 65
DB 146 HATDADSRARASTSAGQAASSAQSSAGTASTKATERSAAAESSKSAATASAGA 205
QY 66 LISLRDAIINKNSPTDLSOLEASTSTSTVTVRAAKDYDEAKSNEDTKSGLENAKTIA 125
DB 206 A-----KTSETNAAVSQSSAATASATATTKAS-----EAAASARDAASSEAKS 250
QY 126 EYETKADIMAAIOMERLANSPSNHTEEVNNIKKALEAKQDITDKLKVTLONOK 185
DB 251 -SETSAAS-----SASASASATPAAGNSAKAA----- 276
QY 186 SLEVEVKTETSDAQIPAINSOLEINKNSADQITKLERONISYEAIVLTNAGEYIKASSEA 245
DB 277 -----KTSE-----TNKSSSETAAEQ-----SASAA 298
QY 246 GIKLQALQSIYDADQSOAAVLAQOONNSPDNIATKELIDAETKVELKOEHTGLTD 305
DB 299 GSKTAAALASASASTSAGQASASATPAAGKSAEASAS--ASTATTKAGATBQASAAAS 355
QY 306 SPLVKAEBQISQAKDIDQIKRSGSDIPVGPSSGAASAGSAGKSSNNSGRISLL 365
DB 356 SASAATSETMAKA---SETSAESSKTAASASSASSASSASASK----- 399
QY 366 DVDNEMAAIALQGFMSIEGFVNPNPATKELQAMEAOLTAASDQLVGADGELPAEIOA 425
DB 400 DEATRQASA-----AKSSATTASTKATE-----A 423
QY 426 IKDALAQAALKQPSADGLATAMGVAPAAKVGGSAG-AGATVQVMNVKQLYKTAFFSSTSS 484
DB 424 ACSATPAASQSKSTAESAAETPAARADIASAVALEDASTTKGIYOL--SSATNSNS 481
QY 485 SSSAAALSGCYGKYKTLNLSY--SESRSGYQASISQTAANALSSVSRSGIESGRADAS 543
DB 482 ESTLAATPKAVKAYELANQYTAQDATTQAKGIYQLSNMTNSTSEMLA-----ATPKS 534
QY 544 QRAAETIVDSQTLGD-----VYSRLQVLDISMSTIYSPQA 580
DB 535 VRAAYDLANGKTYTAQDATTQAKGIYQLSSATNSASSETLATPKA 578

RESULT 36
S02771
myosin heavy chain A [similarity] - Caenorhabditis elegans
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Caenorhabditis elegans
C:Date: 31-Dec-1993 #sequence_revision 19-May-2000 #text_change 19-Apr-2002
C:Accession: T23622; S02771
R:Harris, B.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19773

A:Accession: T23622
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1992 <WLI>
A:Cross-references: EMBL:Z78199; PIDN:CAB01576.1; GSPDB:GN00023; CESP:K12F2.1
R:DiDio, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.
J. Mol. Biol. 205, 603-613, 1989
A:Title: Sequence analysis of the complete Caenorhabditis elegans myosin heavy chain
A:Reference number: S02771; MUID:89178677; PMID:2926820
A:Accession: S02771
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1116,140-1992 <DIB>
A:Cross-references: EMBL:X08067; NID:g6798; PIDN:CAA30856.1; PID:g6799
C:Genetics:
A:Gene: myo-3; CESP:K12F2.1
A:Map position: 5
A:Introns: 46/1; 192/1; 292/1; 468/2; 1921/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle
F:89-802/Domain: myosin motor domain homology <MMOT>
F:202-209/Region: nucleotide-binding motif A (P-loop)
F:690-712/Region: actin binding #status predicted
F:793-807/Region: actin binding #status predicted
F:875-1189/Domain: coiled coil #status predicted <COI>
F:1190-1192/Region: light meromyosin
F:153/Modified site: N6/N6-trimethyllysine (Lys) #status predicted
F:208/Binding site: ATP (Lys) #status predicted
F:730/740/Active site: Cys #status predicted

Query Match 6.2%; Score 200; DB 1; Length 1992;
Best Local Similarity 18.9%; Pred. No. 0.84;
Matches 142; Conservative 152; Mismatches 304; Indels 154; Gaps 28;

QY 28 GQIASSETKSTKASESPASSSSVSM-SFLSSAKNAL---ISLRDAIINKNSPTDS 83
DB 886 GELAVYKIQLE--EAVQREIARSOLESOVADLVEKNAFLSTETKRNLAADAEERKE 943
QY 84 LSOLEASTSTVTVRAADYDEAKSNFDTAKSGLENAKTIAEYETKMAIDMAALODMER 143
DB 944 LNDLKA-TLESKLSDTGLEDQMERNDLARKKTTDOELSDTKKHVDDLSLRKAQ 1002
QY 144 -----LANSDP-----NNHTEEVN-NIKKALEAKQDITDKLNLK-- 177
DB 1003 EKOSRDHNIRSLQDEKANDQEAVALNKREKKHQQEENRKLINEDQSEEDKVHLEKIRNK 1062
QY 178 -----VTIQONKSLTEYK-----TTSDADDIPA----- 202
DB 1063 LEQMDLEENIDREKRSGLIEKARKVEGDLKVAQENIDETTRQKHDEVTTLRKEED 1122
QY 203 ---INSOLEINKN---SAQOITKLERONISYEAVL---TNAGEYIKASSAGIKLQGA 252
DB 1123 LHHTNAKLAENNSITAKLORLIKELTARNAEIEBELEARNRSRQSDRSRAERLEEL 1182
QY 253 LQSIYDADQSOAAVLAQOONNSPDNIATKELIDAA--ETKVELAQE---TGLTD 305
DB 1183 TERLEQOGG-ATAAQLAANKKREAEIARKREKEEDSLNHETALISLRKHGDSVAELTE 1241
QY 306 -----SPLVKAEBQISQAKDIDQ-----IKPSGSDITVGPS 339
DB 1242 QLETLOLKAKESEVSKLQDRLESDQHATDSEVSRQDLEKALRTIEQYSELQTKADE 1301
QY 340 GSAASAGSAAQALKSSNNSGRISLLDDVDNEMAAI--ALQGFMSIEGFVNPN----- 391
DB 1302 QSRQLODFPALNRLNLENSDLNRSLEEDNQLNSLHRKSTLQQLDDETRNAYDESE 1361
QY 392 ----PATAKELQAMEAOLTAASDQLVGADGELPAEIOAIKDALQALQKPSADGLATANG 447
DB 1362 RQALATATAKLEHENTILNEHLEDEAKSKADLTTRQISKMAETIQQMKARFQSEGL-NKLE 1420
QY 448 QVAPAAKVGGSAGTAGTAVQNMNVKQLYKTAFFSSTSSSYAAALSDGYSAVYKTLNLSYSE 507

```

Db 1421 EIE-AKK-----ALQIKVDEL-----TDINEGLFAKIASOEKVRFKIMQDL-DD 1463
QY 508 SSGVOSALISOTANPALSSVSSSGIESGGRSADASQRAETIVRSQTLG-DVYSRLQY 566
Db 1464 AOSDVEKAAQAQVFEKHHRRQESITIAEKKKTKTDLSSELDAQAQRNROLSTDLFRAKTA 1523
QY 567 LDSLMTIYSNPOANE--EIMOKLTASISKAPOFGYPVAVONSADSLQKFAQLEREFPVD 624
Db 1524 NDLAEYLDSTREKNSLAQEVKDLTDQLEGGR-----SVALLQIKVRLKEVEKEE 1575
QY 625 GERSLAESQENAFKQPAFIQOVLVNIASLFS 656
Db 1576 LQKALDEA-EAALAEAEKAVLRQAEVSQIRS 1606

RESULT 37
C97038
phage-related protein, yqpo B, subtilis homolog [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: C97038
R:Noilling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clc
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: C97038
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-2052 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79094.1; PID:q15024039; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
A:Genetics:
A:Gene: CAC1120

Query Match 6.2%; Score 200; DB 2; Length 2052;
Best Local Similarity 18.8%; Pred. No. 0.88;
Matches 137; Conservative 136; Mismatches 289; Indels 166; Gaps 26;

QY 26 INQIANSSEKSTKASESPASSVSSWFLSSAKNALISLRDAIINKNSPDSIS 85
Db 743 LRKAVAGTMTBE--EASALEAMQIKM-----KATVSAEBA-----SSANNALS 786
QY 86 QLEASSTSTV-----TVAAKD-----YDEKSNPDFAKSGLENAKTLA 125
Db 787 VAKRSATLSTGCLNALGSIQVAGLTJGLSMALSDIIEKHEDKQKVDKLQGYE----- 840
QY 126 EYETKADIMAAIQMERLANSDPSSNNHTEEVNNIKKALEAQKDTIDKINKLVTTLQONK 185
Db 841 -----DLTKAMKONSITADNDYKNLEKEQSTLENALKRKALEGIKINSVSKSKKDK 893
QY 186 ---SLTEVTKTDSADQIPAINSOLEINKNSAD-----QIIRKLEKONISYE 229
Db 894 WGNSTLDDTKLDFOSELKNVKNVSETQKNLDTGTGYSDKTTGKIKLSEASEQJANNK 953
QY 230 AVLTVNAGEVTKASSEAGIKTIGALQSTVDAGQSQAAVLAQAQNNNS-----PDNIA 280
Db 954 STEGVARDLTENSDDKTKTIGQLIQSYVTLKYOSDKTALSEQOKLAQALALVFQDLTL 1013
QY 281 ATKRELIDAAETRYNEL-KOEHTGLTDSPLVKKA---EOIQAQKNDIOETIKPSGSDIPV 336
Db 1014 STDKNGDTIIRKNASALQKQTEAFIKGNNAVVAAILDKDELQQAQEKINK----- 1062
QY 337 GPGSGAASAGSAGALKSSNNSGRISLLDD-----VDNEKAIALQGFNSMIEQFN 389
Db 1063 -----AIQSTTNSKYSLYSNOIERNNKIDAOQSVADIKRKHDELVPYLLNN 1108
QY 390 NNPARAKELQAMEA---QLTAMS-DQLVGAD---GELPAEIOAKIALQALAK-OPSADG 441
Db 1109 TIKQLISKIQGGISSFKNTYVVMYAQMOMAEOKMYSKLGEIYQYTYLKQKQISKNNKSSKE 1168
QY 442 LATAMQVAFAPAA---KVGGSGAGT-----AGTVQMMNV-----KQLYKT 477

```

```

Db 1169 LADAKQVOKIKIADLANIKGLSTVTDKNGNTTLKAGNKEKEIKLIKSEGNTRYKELEKI 1228
QY 478 AFSSTSSSSYAALSDGYA-----KTLNLSYSESRCQ---SALISQANANA 523
Db 1229 KUGSSKSTT---MIEGNTAVTVOKQAQYRIGVYTKAKMIAAKAKVYINASTLEIKAGQA 1285
QY 524 LRSVRSRGIESGGRSADASQRAETIVRSQTLGDPVSRLO-VLDSLMTIYSNPOAN- 581
Db 1286 LTSKAKNAKE-----RIKALQEIIDLKASKEIDKLFQADQSADSSNNTVDSQDIGY 1341
QY 582 -OEIMOKLTASISKAPOFGYPVAVONSADSLQKFAQLEREFPVDGERSLAESQENAFRQ 640
Db 1342 ASEDAERKAERKAERKAERKAERKAERKAERKAERKAERKAERKAERKAERKAERKAER 1401
QY 641 PAFIQOVL 648
Db 1402 KKHMEAL 1409

RESULT 38
S45781
Probable calcium-binding protein YBL047c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBL0520
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Apr-2002
C:Accession: S45781; S50284; S45782; S39841; S37339; S42498
R:Goffeau, A.; Jonniaux, J.L.; Purnelle, B.; Skala, J.; de Wergifosse, P.; van Dyck,
A:Reference number: S45745
A:Reference number: S45781
A:Molecule type: DNA
A:Residues: 1-961 <GOF>
A:Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBL047c
A:Experimental source: strain S288C
R:de Wergifosse, P.; Jacques, B.; Jonniaux, J.L.; Purnelle, B.; Skala, J.; Goffeau, A
Yeast 10, 1489-1496, 1994
A:Title: The sequence of a 22.4 kb DNA fragment from the left arm of yeast chromosome
NM-binding protein.
A:Reference number: S50284; MUID:95176707; PMID:7871888
A:Accession: S50284
A:Molecule type: DNA
A:Residues: 1-961 <DEF>
A:Cross-references: EMBL:X78214
A:Experimental source: strain S288C
R:Dubois, E.; El Bakkoury, M.; Glandsdorff, N.; Messenguy, F.; Pierard, A.; Scherens,
A:Reference number: S45782
A:Accession: S45782
A:Molecule type: DNA
A:Residues: 579-1381 <DUB>
A:Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBL047c
A:Experimental source: strain S288C
R:Scherens, B.; el Bakkoury, M.; Vierendeels, F.; Dubois, E.; Messenguy, F.
Yeast 9, 1355-1371, 1993
A:Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of
A:Reference number: S39824; MUID:94205266; PMID:8154187
A:Accession: S39841
A:Molecule type: DNA
A:Residues: 579-1381 <SCH>
A:Cross-references: EMBL:Z23261; MID:g313733; PIDN:CAA80797.1; PID:g313748
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:EDI1; MIPS:YBL047c
A:Cross-references: SGD:S0000143
A:Map position: 2L
C:Superfamily: yeast probable calcium-binding protein YBL047c; calmodulin repeat homo
C:Keywords: calcium binding; EF hand; transmembrane protein
F:167-199/Domain: calmodulin repeat homology <EPI>
F:560-576/Domain: transmembrane #status predicted <TMM>

Query Match 6.2%; Score 199.5; DB 1; Length 1381;
Best Local Similarity 20.4%; Pred. No. 0.54;

```


THIS PAGE BLANK (USPTO)